

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|  |           |   |
|--|-----------|---|
| <b>(51) International Patent Classification <sup>5</sup>:</b><br>A61K 39/12, 39/00, C07K 17/00, 3/00,<br>13/00, 15/00, C07H 15/12, 13/00   | <b>A1</b> | <b>(11) International Publication Number:</b> WO 94/22477<br><b>(43) International Publication Date:</b> 13 October 1994 (13.10.94)   |
| <b>(21) International Application Number:</b> PCT/US94/03282<br><b>(22) International Filing Date:</b> 25 March 1994 (25.03.94)<br><b>(30) Priority Data:</b><br>08/037,816 26 March 1993 (26.03.93) US<br><b>(60) Parent Application or Grant</b><br><b>(63) Related by Continuation</b><br>US 08/037,816 (CIP)<br>Filed on 26 March 1993 (26.03.93)<br><b>(71) Applicant (for all designated States except US):</b> PROGENICS<br>PHARMACEUTICALS, INC. [US/US]; Old Saw Mill River<br>Road, Tarrytown, NY 10591 (US).<br><b>(72) Inventors; and</b><br><b>(75) Inventors/Applicants (for US only):</b> HASEL, Karl, W.<br>[CH/US]; 1713 Baldwin Road, Yorktown Heights, NY<br>10598 (US). MADDON, Paul, J. [US/US]; Apartment<br>#25C, 60 Haven Avenue, New York, NY 10032 (US).<br><b>(74) Agent:</b> WHITE, John, P.; Cooper & Dunham, 30 Rockefeller<br>Plaza, New York, NY 10112 (US). |           | <b>(81) Designated States:</b> AU, BR, CA, CN, JP, RU, UA, US,<br>European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR,<br>IE, IT, LU, MC, NL, PT, SE).<br><br><b>Published</b><br>With international search report. |
| <b>(54) Title:</b> HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED THERETO, AND THERAPEUTIC AND PROPHYLACTIC<br>USES THEREOF   |           |   |
| <b>(57) Abstract</b><br><br>The invention provides a recombinant nucleic acid molecule which encodes a mutant HIV-1 gp120 envelope glycoprotein, vaccines comprising the mutant HIV-1 envelope glycoprotein, antibodies and methods of treating individuals.   |           |   |

BEST AVAILABLE COPY

Applicants: G.P. Allaway, et al.  
Serial No.: 09/724,105  
Filed: November 28, 2000  
Exhibit 1

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |  |    |                          |
|----|--------------------------|----|--|----|--------------------------|
| AT | Austria                  | GB | United Kingdom                           | MR | Mauritania               |
| AU | Australia                | GE | Georgia                                  | MW | Malawi                   |
| BB | Barbados                 | GN | Guinea                                   | NE | Niger                    |
| BE | Belgium                  | GR | Greece                                   | NL | Netherlands              |
| BF | Burkina Faso             | HU | Hungary                                  | NO | Norway                   |
| BG | Bulgaria                 | IE | Ireland                                  | NZ | New Zealand              |
| BJ | Benin                    | IT | Italy                                    | PL | Poland                   |
| BR | Brazil                   | JP | Japan                                    | PT | Portugal                 |
| BY | Belarus                  | KE | Kenya                                    | RO | Romania                  |
| CA | Canada                   | KG | Kyrgyzstan                               | RU | Russian Federation       |
| CF | Central African Republic | KP | Democratic People's Republic<br>of Korea | SD | Sudan                    |
| CG | Congo                    | KR | Republic of Korea                        | SE | Sweden                   |
| CH | Switzerland              | KZ | Kazakhstan                               | SI | Slovenia                 |
| CI | Côte d'Ivoire            | LI | Liechtenstein                            | SK | Slovakia                 |
| CM | Cameroon                 | LK | Sri Lanka                                | SN | Senegal                  |
| CN | China                    | LU | Luxembourg                               | TD | Chad                     |
| CS | Czechoslovakia           | LV | Latvia                                   | TG | Togo                     |
| CZ | Czech Republic           | MC | Monaco                                   | TJ | Tajikistan               |
| DE | Germany                  | MD | Republic of Moldova                      | TT | Trinidad and Tobago      |
| DK | Denmark                  | MG | Madagascar                               | UA | Ukraine                  |
| ES | Spain                    | ML | Mali                                     | US | United States of America |
| FI | Finland                  | MN | Mongolia                                 | UZ | Uzbekistan               |
| FR | France                   |    |  | VN | Viet Nam                 |
| GA | Gabon                    |    |  |    |                          |

5 HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED THERETO, AND  
THERAPEUTIC AND PROPHYLACTIC USES THEREOF

10 Background of the Invention

10 Throughout this application, various publications are  
referenced by Arabic numerals. Full citations for these  
references may be found at the end of the specification  
immediately preceding the claims. The disclosure of these  
15 publications is hereby incorporated by reference into this  
application to describe more fully the art to which this  
invention pertains.

20 The life cycle of animal viruses is characterized by a  
series of events that are required for the productive  
infection of the host cell. The initial step in the  
replicative cycle is the attachment of the virus to the cell  
surface, which attachment is mediated by the specific  
interaction of the viral attachment protein (VAP) to  
25 receptors on the surface of the target cell. The  
differential pattern of expression of these receptors is  
largely responsible for the host range and tropic properties  
of viruses. In addition, an effective immune response  
against many viruses is mediated through neutralizing  
30 antibodies directed against the VAP. The interaction of the  
VAP with cellular receptors and the immune system therefore  
plays a critical role in infection and pathogenesis of viral  
disease.

35 The human immunodeficiency virus type 1 (HIV-1) infects  
primarily helper T lymphocytes, dendritic cells, and  
monocytes/macrophages--cells that express surface CD4--  
leading to a gradual loss of immune function. This loss of  
function results in the development of the human acquired

immunodeficiency syndrome (AIDS) (1). The initial phase of the HIV-1 replicative cycle involves the high-affinity interaction between the HIV-1 exterior envelope glycoprotein gp120 and cell surface CD4 ( $K_d$  approximately  $4 \times 10^{-9}$  M) (2).  
5 Several lines of evidence demonstrate the requirement of this interaction for viral infectivity. The introduction into CD4<sup>+</sup> human cells of cDNA encoding CD4 is sufficient to render otherwise resistant cells susceptible to HIV-1 infection (3). In vivo, viral infection appears to be  
10 restricted to cells expressing CD4, indicating that the cellular tropism of HIV-1 is largely determined by the pattern of cellular expression of CD4. Following the binding of HIV-1 gp120 to cell surface CD4, viral and target cell membranes fuse by a mechanism that is poorly  
15 understood, resulting in the introduction of the viral capsid into the target cell cytoplasm (4).

Mature CD4 has a relative molecular mass ( $M_r$ ) of 55 kDa and consists of an N-terminal 372-amino acid extracellular  
20 domain containing four tandem immunoglobulin-like regions (V1-V4), followed by a 23-amino acid transmembrane domain and a 38-amino acid cytoplasmic segment (5, 6). In experiments using truncated sCD4 proteins, it has been shown that the determinants for high-affinity binding to HIV-1  
25 gp120 lie solely within the N-terminal immunoglobulin-like domain (V1) (7-9). Mutational analysis of V1 has defined a discrete binding site (residues 38-52) that comprises a region structurally homologous to the second complementarity-determining region (CDR2) of immunoglobulin  
30 genes (9).

The production of large quantities of sCD4 has permitted a structural analysis of the two N-terminal immunoglobulin-like domains (V1V2). The structure determined at 2.3



angstrom resolution reveals that the molecule has two tightly-associated domains, each of which contains the immunoglobulin-fold connected by a continuous beta strand. The putative binding sites for monoclonal antibodies, class II major histocompatibility complex (MHC) molecules, and HIV-1 gp120, as determined by mutational analyses, map on the molecular surface (10, 11).

The HIV-1 envelope gene env encodes an envelope glycoprotein precursor, gp160, which is cleaved by cellular proteases before transport to the plasma membrane to yield gp120 and gp41. The membrane-spanning glycoprotein, gp41, is non-covalently associated with gp120, a purely extracellular glycoprotein. The mature gp120 molecule is heavily glycosylated (approximately 24 N-linked oligosaccharides), contains approximately 480 amino acid residues with 9 intra-chain disulfide bonds (12), and projects from the viral membrane as a dimeric or multimeric molecule (13).

Mutational studies of HIV-1 gp120 have delineated important functional regions of the molecule. The regions of gp120 that interact with gp41 map primarily to the N- and C-termini (14). The predominant strain-specific neutralizing epitope on gp120 is located in the 32-34 amino acid residue third variable loop, herein referred to as the V3 loop, which resides near the center of the gp120 sequence (15). The CD4 binding site maps to discontinuous regions of gp120 that include highly conserved or invariant amino acid residues in the second, third, and fourth conserved domains (the C2, C3, and C4 domains) of gp120 (16). It has been postulated that a small pocket formed by these conserved residues within gp120 could accommodate the CDR2 loop of CD4, a region defined by mutational analyses as important in interacting with gp120 (17).

HIV-1 gp120 not only mediates viral attachment to surface CD4 molecules, but also serves as the major target of antibodies which neutralize non-cell-associated virus and inhibit cell to cell viral transmission.

5

There are two major classifications of HIV-1-neutralizing antibodies: type-specific and group-common (15). Type-specific neutralizing antibodies primarily recognize linear determinants in the highly variable V3 loop of gp120. These antibodies act by inhibiting fusion between HIV-1 and the target cell membrane, and generally neutralize only a particular isolate of, or closely related strains of, HIV-1. Sequence variation within the V3 loop, as well as outside of this region, permits viruses to escape neutralization by anti-V3 loop antibodies. In contrast, group-common neutralizing antibodies primarily recognize discontinuous or conformational epitopes in gp120, and possess the ability to neutralize a diverse range of HIV-1 isolates. These broadly neutralizing antibodies often recognize a site on gp120 which overlaps the highly conserved CD4-binding site, and thus inhibits gp120-CD4 binding.

A structural relationship has been demonstrated between the V3 loop and the C4 region of gp120 which region constitutes both part of the CD4 binding site and part of the conserved neutralization epitopes. It was observed that deleting the V3 loop resulted in significantly increased binding of a panel of broadly neutralizing hMoAbs (neutralizing human monoclonal antibodies) to the CD4 binding site (18).

30

A major goal in AIDS vaccine development is to develop a vaccine able to protect a subject against the numerous genetic variants of HIV-1 that infect humans. Although cell-mediated immune responses might serve to control infection in HIV-1-infected individuals, several lines of

35

evidence demonstrate that protection against infection is mainly mediated by neutralizing antibodies directed against gp120. Early experiments showed that immunization of chimpanzees with recombinant gp120 induced a protective immune response against challenge with the homologous HIV-1 strain (17). This protection correlated with the presence of high-titer neutralizing antibodies against the V3 loop of gp120. In addition, passive immunization of chimpanzees with a V3-loop neutralizing monoclonal antibody resulted in protection against challenge with the homologous HIV-1 strain (19). Although protection against challenge was demonstrated in these two experiments, recent studies have questioned the clinical relevance of these findings. For example, these neutralizing antibodies recognize the V3 loop determinants of a single strain, and not conserved or discontinuous epitopes. Thus, these antibodies lack the ability to neutralize the broad spectrum of HIV-1 strains present in an HIV-1 population. Furthermore, the challenge virus was the homologous HIV-1 laboratory adapted LAI (HTLV-IIIB) strain and not one of the primary isolates that contain considerable gp120 sequence heterogeneity. Since these experiments showed that gp120 subunit vaccination induces an immune response effective against only the homogeneous HIV-1 strain used as an antigen, it is unlikely that the vaccination regimens used in these studies would be useful in humans.

Individuals infected by HIV-1 typically develop antibodies that neutralize the virus in vitro, and neutralization titers decrease with disease progression (19). Analysis of sera from HIV-1-infected humans indicates that type-specific neutralizing antibodies appear early in infection. Later in the course of infection, a more broadly neutralizing antibody response develops. However this antibody response is of significantly lower titer and/or affinity.

Fractionation studies of HIV-1 antibody-positive human sera reveal that the type-specific neutralizing activity is primarily directed against linear determinants in the V3 loop of gp120 (20). There was no correlation found among  
5 antibodies between the ability to neutralize divergent HIV-1 isolates and reactivity to the V3 loop of these isolates. In contrast, the broadly neutralizing antibodies present in HIV-1 antibody-positive human sera primarily recognize discontinuous epitopes in gp120 which overlap the CD4-  
10 binding site and block gp120-CD4 binding. In other words, the broadly neutralizing activity of neutralizing antibodies is not merely the result of additive anti-V3 loop reactivities against diverse HIV-1 isolates which appear during infection.

15 Recently, several groups have generated human monoclonal antibodies (hMoAbs) derived from HIV-1 infected individuals which possess type-specific or group-common neutralizing activities (17). The type-specific neutralizing hMoAbs were  
20 found to recognize linear determinants in the V3 loop of gp120. In contrast, the group-common neutralizing hMoAbs generally recognize discontinuous epitopes which overlap the CD4-binding site and block gp120-CD4 binding.

25 The V3 loop is a highly immunodominant region of gp120 which partially interacts with the CD4-binding region. The presence of the V3 loop region on gp120 may skew the humoral immune response away from producing antibodies which specifically bind to the CD4-binding domain of gp120.  
30 Furthermore, the advantages of removing the V3 loop to expose the CD4-binding domain of gp120 to the immune system would be countered by the fact that the exposed CD4-binding site would still have a high affinity for cell surface CD4. In other words, a mutant gp120 protein missing only the V3  
35 loop would quickly bind to CD4+ cells and would thus be

hampered in generating an immune response against the exposed CD4-binding site.

The subject invention provides a mutant HIV-1 gp120 envelope glycoprotein which overcomes both the problems of V3 loop immunodominance and of the high affinity to CD4. The subject invention further provides vaccines comprising the mutant HIV-1 gp120 envelope glycoprotein, antibodies which specifically bind to the CD4-binding site of HIV-1 gp120 envelope glycoprotein, pharmaceutical compositions comprising these antibodies, and methods of using these vaccines and compositions to treat or prevent HIV-1 infection.

Summary of the Invention

The subject invention provides a recombinant nucleic acid molecule which encodes a mutant HIV-1 gp120 envelope glycoprotein comprising a V3 loop deletion and a C4 domain<sub>(W-X)</sub> point mutation, wherein X is an amino acid residue other than tryptophan. In the preferred embodiment, X is a valine residue.

10 In one embodiment, the nucleic acid molecule is a DNA molecule. The DNA molecule may be a plasmid. In one embodiment, the plasmid comprises the sequence of the plasmid designated PPI4-tPA.

15 In one embodiment, the C4 domain is an HIV-1<sub>LA1</sub> gp120 envelope glycoprotein C4 domain. The mutant HIV-1 gp120 envelope glycoprotein may be a mutant HIV-1<sub>LA1</sub> gp120 envelope glycoprotein.

20 In another embodiment, the C4 domain is an HIV-1<sub>JR-FL</sub> gp120 envelope glycoprotein C4 domain. The mutant HIV-1 gp120 envelope glycoprotein may be a mutant HIV-1<sub>JR-FL</sub> gp120 envelope glycoprotein.

25 The subject invention also provides the mutant HIV-1 gp120 envelope glycoprotein encoded by the recombinant nucleic acid molecule of the subject invention.

The subject invention further provides a vaccine which  
30 comprises a therapeutically effective amount of the mutant HIV-1 gp120 envelope glycoprotein of the subject invention, and an adjuvant.

The subject invention further provides a method of treating

an HIV-1-infected subject, which comprises immunizing the HIV-1-infected subject with the vaccine of the subject invention, thereby treating the HIV-1-infected subject.

- 5 The subject invention further provides a vaccine which comprises a prophylactically effective amount of the mutant HIV-1 gp120 envelope glycoprotein of the subject invention, and an adjuvant.
- 10 The subject invention further provides a method of reducing the likelihood of an HIV-1-exposed subject's becoming infected with HIV-1, which comprises immunizing the HIV-1-exposed subject with the vaccine of the subject invention, thereby reducing the likelihood of the HIV-1-exposed
- 15 subject's becoming infected with HIV-1.

- The subject invention further provides a method of reducing the likelihood of a non-HIV-1-exposed subject's becoming infected with HIV-1, which comprises immunizing the non-HIV-
- 20 1-exposed subject with the vaccine of the subject invention, thereby reducing the likelihood of the non-HIV-1-exposed subject's becoming infected with HIV-1.

- The subject invention further provides a method of obtaining
- 25 partially purified antibodies which specifically bind to the CD4-binding domain of HIV-1 gp120 envelope glycoprotein, which method comprises (a) immunizing a non-HIV-1-exposed subject with the vaccine of the subject invention, (b) recovering from the immunized subject serum comprising said
- 30 antibodies, and (c) partially purifying said antibodies, thereby obtaining partially purified antibodies which specifically bind to the CD4-binding domain of HIV-1 gp120 envelope glycoprotein. In the preferred embodiment, the subject is a human.

In one embodiment, the subject is a medical practitioner.  
In another embodiment, the subject is a newborn infant.

- 5 Finally, the subject invention provides a method of reducing  
the likelihood of a non-HIV-1-exposed subject's becoming  
infected with HIV-1 as a result of exposure thereto during  
an incident wherein there is an increased risk of exposure  
to HIV-1, which comprises administering to the subject  
10 immediately prior to the incident a dose of the composition  
of the subject invention effective to reduce the population  
of HIV-1 to which the subject is exposed during the  
incident, thereby reducing the likelihood of the subject's  
becoming infected with HIV-1. In one embodiment, the  
15 subject is a medical practitioner.



Brief Description of the FiguresFigure 1

gp120 structure. Shown is a box diagram of HIV-1 gp120 depicting the boundaries of the five constant domains (C1-C5) and the five variable domains (V1-V5). The amino acid residue numbering above the box begins at the initiator methionine found at the beginning of the signal sequence (S) and is approximated based on a consensus of all known HIV-1 gp120 amino acid sequences. Also shown are the C4 domain amino acid sequences of HIV-1 strains LAI and JR-FL. Above the C4 domain sequences are indicated two mutations that reduce gp120 binding to cell surface CD4; tryptophan to valine and aspartate to alanine.

15

Figure 2

PPI4-tPA-gp120<sub>LAI</sub>. Expression vector with the HIV-1<sub>LAI</sub> gp120 gene fused to the CMV MIE promoter, and the tPA signal sequence replacing the HIV-1 gp120 signal sequence. Abbreviations: CMV MIE = cytomegalovirus major immediate early, E = enhancer, P = promoter, EXA = Exon A, INA = Intron A, EXB = Exon B, tPA ss = human tissue plasminogen activator signal sequence, gp120 = glycoprotein 120, BGH = bovine growth hormone, AMP = ampicillin resistance gene, and DHFR = dihydrofolate reductase gene.

25

Figure 3

CMV MIE promoter fused to tPA-gp120<sub>LAI</sub>. The nucleotide sequence of the CMV MIE promoter/enhancer region is shown fused to the HIV-1<sub>LAI</sub> gp120 gene that contains the tPA signal sequence. The numbering of nucleotide sequence begins with the HincII site and the numbering of the amino acid sequence begins with the first methionine found in the tPA signal sequence. The tPA signal sequence is fused in-frame to Thr<sub>31</sub>.

30

of gp120, the first amino acid found in mature gp120. The signal sequence is shown in bold as are various landmark restriction sites used for cloning as discussed in the text. The locations of Exon A, Intron A, Exon B and the transcription start site and the signal cleavage site are indicated.

#### Figure 4

Transient expression of gp120. Autoradiograph of <sup>35</sup>S-labeled supernatants from COS cell transfectants, immunoprecipitated with a CD4-immunoglobulin-Protein A-Sepharose complex, and run on a reducing 10% SDS-PAGE gel. The plasmids used for transfection were: Lane 1: Mock transfected cells; lane 2: a vector encoding a CD4-immunoglobulin chimera as a positive transfection control; lane 3: PPI4-tPA-gp120<sub>LAI</sub>; and lane 4: PPI4-tPA-gp120<sub>JR-FL</sub>. Positions of molecular weight markers are indicated.

#### Figure 5

Determination of gp120 concentration by ELISA. Panel A: Concentrations of gp120 in media of CHO cell lines, stably transfected with PPI4-tPA-gp120<sub>LAI</sub>, determined by ELISA. Panel B: A standard curve was established using known amounts of gp120.

#### Figure 6

Expression of gp120 in stably transfected CHO cells. Autoradiograph of <sup>35</sup>S-labeled supernatants from stable CHO cell lines, immunoprecipitated with a CD4-immunoglobulin-Protein A-Sepharose complex, and run on a reducing 10% SDS-PAGE gel. Lane 1: clone 9; lane 2: clone 13; lane 3: clone 6; lane 4: Clone 5. Positions of molecular weight markers are indicated.

Figure 7

tPA-gp120<sub>JR-FL</sub>. The nucleotide and deduced amino acid sequence of the tPA signal sequence fused to HIV-1<sub>JR-FL</sub> gp120 is shown. The NarI and NotI restriction endonuclease sites used for cloning are shown in bold. The predicted site of cleavage by signal peptidase between Arg<sub>35</sub> and Val<sub>36</sub> is indicated.

Figure 8

tPA-gp120<sub>LAI</sub>-V3<sup>(c)</sup>. The nucleotide and deduced amino acid sequence of the tPA signal sequence fused to HIV-1<sub>LAI</sub> gp120 with the V3 loop deleted and replaced with the pentapeptide TGAGH is shown. The V3 loop replacement and the NarI and NotI restriction endonuclease sites used for cloning are shown in bold. The predicted site of cleavage by signal peptidase between Arg<sub>35</sub> and Thr<sub>36</sub> is indicated.

Figure 9

tPA-gp120<sub>JR-FL</sub>-V3<sup>(c)</sup>. The nucleotide and deduced amino acid sequence of the tPA signal sequence fused to HIV-1<sub>JR-FL</sub> gp120 with the V3 loop deleted and replaced with the pentapeptide TGAGH is shown. The V3 loop replacement and the NarI and NotI restriction endonuclease sites used for cloning are shown in bold. The predicted site of cleavage by signal peptidase between Arg<sub>35</sub> and Val<sub>36</sub> is indicated.

Figure 10

tPA-gp120<sub>LAI</sub>-V3<sup>(c)</sup>-CD4<sup>(c)</sup>. Shown is the nucleotide and deduced amino acid sequence of the tPA signal sequence fused to HIV-1<sub>LAI</sub> gp120, with the V3 loop deleted and replaced with the pentapeptide TGAGH, and Trp<sub>408</sub> mutated to Val. The mutations and the NarI and NotI restriction endonuclease sites used for cloning are shown in bold. The predicted site of cleavage by signal peptidase between Arg<sub>35</sub> and Thr<sub>36</sub> is indicated.

Figure 11

tPA-gp120<sub>JR-FL</sub>-V3<sup>(-)</sup>-CD4<sup>(-)</sup>. Shown is the nucleotide and deduced amino acid sequence of the tPA signal sequence fused to HIV-1<sub>JR-FL</sub> gp120, with the V3 loop deleted and replaced with the pentapeptide TGAGH, and Trp<sub>396</sub> mutated to Val. The mutations and the NarI and NotI restriction endonuclease sites used for cloning are shown in bold. The predicted site of cleavage by signal peptidase between Arg<sub>35</sub> and Val<sub>36</sub> is indicated.

10

Figure 12

tPA-gp120<sub>LAI</sub>-CD4<sup>(-)</sup>. Shown is the nucleotide and deduced amino acid sequence of the tPA signal sequence fused to HIV-1<sub>LAI</sub> gp120. The Trp<sub>437</sub> to Val CD4 binding mutation, the NarI and NotI restriction endonuclease sites used for cloning, and the predicted site of cleavage by signal peptidase between Arg<sub>35</sub> and Thr<sub>36</sub> are shown in bold.

Figure 13

20 tPA-gp120<sub>JR-FL</sub>-CD4<sup>(-)</sup>. Shown is the nucleotide and deduced amino acid sequence of the tPA signal sequence fused to HIV-1<sub>JR-FL</sub> gp120. The Trp<sub>424</sub> to Val CD4 binding mutation, the NarI and NotI restriction endonuclease sites used for cloning and the predicted cleavage by signal peptidase between Arg<sub>35</sub> and Val<sub>36</sub> are shown in bold.

25

Figure 14Expression of gp120 in stably transfected CHO cells.

Autoradiograph of super <sup>35</sup>S-labeled supernatants from stable CHO cell lines, immunoprecipitated with MoAb F105-Protein A-Sepharose complex, and run on a reducing 10% SDS-PAGE gel. Panel A: Lane 1: tPA-gp120<sub>LAI</sub> CHO cells; lane 2: tPA-gp120<sub>LAI</sub>-V3<sup>(-)</sup> CHO cells; lane 3: tPA-gp120<sub>LAI</sub>-V3<sup>(-)</sup>-CD4<sup>(-)</sup> CHO cells. Panel B: Lane 1: tPA-gp120<sub>JR-FL</sub> CHO cells; lane 2: tPA-gp120<sub>JR-FL</sub>-V3<sup>(-)</sup>

30

CHO cells; lane 3: tPA-gp120<sub>JR-FL</sub>-V3<sup>(-)</sup>-CD4<sup>(-)</sup> CHO cells. Positions of molecular weight markers are indicated.

Figure 15

5 Purified gp120 proteins.

Silver stained 10% SDS-PAGE gel with a sample of purified gp120 proteins. Panel A: Lane 1: tPA-gp120<sub>LAI</sub> CHO cells; lane 2: tPA-gp120<sub>LAI</sub>-V3<sup>(-)</sup> CHO cells; lane 3: tPA-gp120<sub>LAI</sub>-V3<sup>(-)</sup>-CD4<sup>(-)</sup> CHO cells. Panel B: Lane 1: tPA-gp120<sub>JR-FL</sub> CHO cells; lane 2: 10 tPA-gp120<sub>JR-FL</sub>-V3<sup>(-)</sup> CHO cells; lane 3: tPA-gp120<sub>JR-FL</sub>-V3<sup>(-)</sup>-CD4<sup>(-)</sup> CHO cells. Positions of molecular weight markers are indicated.

Figure 16

15 Analysis of binding of recombinant mutant gp120 to cell surface human CD4 by FACS.

Plate 1. DG44 cells, a subclone of CHO cells which lack expression of the human CD4 protein, were used as control. Increasing concentrations of HIV-1 gp120<sub>LAI</sub> did not show an increase in specific fluorescence when compared to 20 background. Plate 2. DG44 #3 cells are a CHO cell line transfected with the cDNA clone encoding the human CD4 protein. Increasing concentrations of HIV-1 gp120<sub>LAI</sub> show a dramatic increase (or shift) in fluorescence. Plate 3. Similar to Plate 2 but the HIV-1 gp120<sub>LAI</sub>-V3<sup>(-)</sup> protein was 25 added. Again a large shift indicating binding to the DG44 #3 cells was seen. Plate 4. DG44 #3 cells were incubated with either HIV-1 gp120<sub>LAI</sub>-V3<sup>(-)</sup>-CD4<sup>(-)</sup> protein or MoAb OKT4A an antibody with high affinity for human CD4. Only OKT4A bound to the cells.

Detailed Description of the Invention

The plasmids designated PPI4-tPA-gp120<sub>LAI</sub> and PPI4-tPA-gp120<sub>JR-FL</sub> were deposited pursuant to, and in satisfaction of, the requirements of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure with the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland 20852 under ATCC Accession Nos. 75431 and 75432, respectively. The plasmids PPI4-tPA-gp120<sub>LAI</sub> and PPI4-tPA-gp120<sub>JR-FL</sub> were deposited with the ATCC on March 12, 1993.

The subject invention provides a recombinant nucleic acid molecule which encodes a mutant HIV-1 gp120 envelope glycoprotein comprising a V3 loop deletion and a C4 domain<sub>(w->x)</sub> point mutation, wherein X is an amino acid residue other than tryptophan. In the preferred embodiment, X is a valine residue.

In one embodiment, the nucleic acid molecule is a DNA molecule. The DNA molecule may be a plasmid. In one embodiment, the plasmid comprises the sequence of the plasmid designated PPI4-tPA.

The V3 loop of HIV-1 gp120 envelope glycoprotein is shown in Figure 1. The V3 loop is demarcated by cysteine residues at both its N- and C-termini. As used herein, a V3 loop deletion means a deletion of one or more amino acid residues between the terminal cysteine residues, with the proviso that there must be three or more amino acid residues situated between the two terminal cysteine residues in a V3 loop deletion. These three or more amino acid residues may either be residues originally present in the V3 loop, or exogenous residues. For example, as shown in the

Experimental Details section *infra*, the pentapeptide TGAGH is situated between the two terminal cysteine residues. Variations in the size of the V3 loop deletion illustrated herein are tolerable without affecting the overall structure of the mutant HIV-1 gp120 envelope glycoprotein, as is well known to those skilled in the art.

As used herein, "C4 domain" means the HIV-1 gp120 envelope glycoprotein C4 domain having the following consensus sequence:

$X_1X_2X_3CX_4IX_5X_6X_7X_8X_9X_{10}\mathbf{W}X_{11}X_{12}X_{13}X_{14}X_{15}AX_{16}YX_{17}X_{18} -$   
 $PX_{19}X_{20}X_{21}X_{22}X_{23}X_{24}X_{25}X_{26}SX_{27}X_{28}TGX_{29}X_{30}X_{31}X_{32}RX_{33}GX_{34},$

wherein  $X_1 = T, I, V, K$  or  $R$ ;  $X_2 = L, I$  or  $H$ ;  $X_3 = P, Q, L$  or  $T$ ;  $X_4 = R, K$  or  $G$ ;  $X_5 = K$  or  $E$ ;  $X_6 = Q$  or  $E$ ;  $X_7 = F, I$  or  $V$ ;  $X_8 = I, V$  or  $M$ ;  $X_9 = N, R$  or  $K$ ;  $X_{10} = M, R, L$  or  $T$ ;  $X_{11} = Q, R$  or  $V$ ;  $X_{12} = E, K, G, R, V$  or  $A$ ;  $X_{13} = V, T, A$  or  $G$ ;  $X_{14} = G$  or  $E$ ;  $X_{15} = K, R, E$ , or  $Q$ ;  $X_{16} = M, V, I$  or  $L$ ;  $X_{17} = A, T$  or  $D$ ;  $X_{18} = P$  or  $L$ ;  $X_{19} = I$  or  $F$ ;  $X_{20} = S, R, G, K, N, A, E$  or  $Q$ ;  $X_{21} = G$  or  $R$ ;  $X_{22} = Q, L, P, N, K, V, T, E$  or  $I$ ;  $X_{23} = I, V$  or  $L$ ;  $X_{24} = R, K, S, N, G, I, T, E$  or  $I$ ;  $X_{25} = C$  or  $R$ ;  $X_{26} = S, L, I, T, P, E, V, K, D$  or  $N$ ;  $X_{27} = N, K$  or  $L$ ;  $X_{28} = I$  or  $V$ ;  $X_{29} = L, P$  or  $I$ ;  $X_{30} = L$  or  $I$ ;  $X_{31} = L$  or  $I$ ;  $X_{32} = T, A, I, V$  or  $E$ ;  $X_{33} = D$  or  $E$ ;  $X_{34} = G$  or  $V$ .

The C4 domain consensus sequence is based on existing C4 domain sequence information from various HIV-1 strains, and thus is not necessarily an exhaustive consensus sequence. The conserved tryptophan residue shown in bold after residue  $X_{10}$  is the only conserved tryptophan residue in the C4 domain. As used herein, a C4 domain<sub>(W->X)</sub> point mutation is a mutation of the above-identified conserved C4 domain tryptophan residue to an amino acid residue other than

tryptophan. For example, a C4 domain<sub>(W->V)</sub> point mutation is a mutation of the conserved C4 domain tryptophan residue to a valine residue.

5 In one embodiment, the C4 domain is an HIV-1<sub>LAI</sub> gp120 envelope glycoprotein C4 domain. The sequence of the HIV-1<sub>LAI</sub> gp120 C4 domain is: TLPCRKQFINMWQEVGKAMYAPPISGQIRCS-SNITGLLLTRDGG. The mutant HIV-1 gp120 envelope glycoprotein may be a mutant HIV-1<sub>LAI</sub> gp120 envelope glycoprotein.

10

In another embodiment, the C4 domain is an HIV-1<sub>JR-FL</sub> gp120 envelope glycoprotein C4 domain. The sequence of the HIV-1<sub>JR-FL</sub> gp120 C4 domain is: 'TLPCRKQIINMWQEVGKAMYAPPIRGQIRCS-SNITGLLLTRDGG. The mutant HIV-1 gp120 envelope glycoprotein  
15 may be a mutant HIV-1<sub>JR-FL</sub> gp120 envelope glycoprotein.

HIV-1<sub>LAI</sub> is a laboratory-adapted strain that is tropic for phytohemagglutinin (PHA)-stimulated peripheral blood lymphocytes (PBLs) and immortalized human T-cell lines. In  
20 contrast, HIV-1<sub>JR-FL</sub> was isolated from brain tissue taken at autopsy that was co-cultured with lectin-activated normal human PBLs. HIV-1<sub>JR-FL</sub> is tropic for PHA-stimulated PBLs and blood-derived macrophages but will not replicate in transformed T-cell lines. Mutant HIV-1 gp120 envelope  
25 glycoproteins derived from a clinical isolate of HIV-1 such as JR-FL may possess new or different epitopes compared to the laboratory-adapted HIV-1 strains that are beneficial for successful vaccination. Although only the HIV-1<sub>LAI</sub> and HIV-1<sub>JR-FL</sub> strains are used herein to generate the mutant HIV-1  
30 gp120 envelope glycoproteins of the subject invention, other HIV-1 strain could be substituted in their place as is well known to those skilled in the art.

The V1 and V2 variable regions of gp120 are unnecessary for



CD4 binding (21). Therefore the mutant HIV-1 gp120 envelope glycoprotein of this invention can either include or exclude the V1 and V2 variable regions.

- 5 The subject invention additionally provides a recombinant nucleic acid molecule which encodes a mutant HIV-1 gp120 envelope glycoprotein comprising a V3 loop deletion and a C4 domain<sub>(Asp→X)</sub> point mutation, wherein the aspartate residue is between amino acid residues X<sub>15</sub> and X<sub>16</sub> in the C4 consensus  
10 sequence, and X is an amino acid residue other than aspartate or glutamate. In the preferred embodiment, X is an alanine residue.

- The subject invention additionally provides a recombinant  
15 nucleic acid molecule which encodes a mutant HIV-1 gp120 envelope glycoprotein comprising a V3 loop deletion and a C4 domain<sub>(Glu→X)</sub> point mutation, wherein the glutamate residue is between amino acid residues X<sub>15</sub> and X<sub>16</sub> in the C4 consensus sequence, and X is an amino acid residue other than  
20 aspartate or glutamate. In the preferred embodiment, X is an alanine residue.

- The subject invention additionally provides a recombinant nucleic acid molecule which encodes a mutant HIV-1<sub>LAI</sub> gp120  
25 envelope glycoprotein comprising a V3 loop deletion and a C3 domain<sub>(asp378→X)</sub> point mutation, wherein X is an amino acid residue other than aspartate or glutamate. In the preferred embodiment, X is a lysine residue.

- 30 The subject invention additionally provides a recombinant nucleic acid molecule which encodes a mutant HIV-1<sub>JR-FL</sub> gp120 envelope glycoprotein comprising a V3 loop deletion and a C3 domain<sub>(asp369→X)</sub> point mutation, wherein X is an amino acid residue other than aspartate or glutamate. In the preferred

embodiment, X is a lysine residue.

The subject invention additionally provides a recombinant nucleic acid molecule which encodes a mutant HIV-1<sub>LAI</sub> gp120 envelope glycoprotein comprising a V3 loop deletion and a C3 domain<sub>(glu380->X)</sub> point mutation, wherein X is an amino acid residue other than glutamate. In the preferred embodiment, X is a glutamine residue.

- 10 The subject invention additionally provides a recombinant nucleic acid molecule which encodes a mutant HIV-1<sub>JR-FL</sub> gp120 envelope glycoprotein comprising a V3 loop deletion and a C3 domain<sub>(glu371->X)</sub> point mutation, wherein X is an amino acid residue other than glutamate. In the preferred embodiment, X is a glutamine residue.

- 20 The subject invention additionally provides a recombinant nucleic acid molecule which encodes a mutant HIV-1<sub>LAI</sub> gp120 envelope glycoprotein comprising a V3 loop deletion and a C2 domain<sub>(thr267->X)</sub> point mutation, wherein X is an amino acid residue other than threonine. In the preferred embodiment, X is an arginine residue.

- 25 The subject invention additionally provides a recombinant nucleic acid molecule which encodes a mutant HIV-1<sub>JR-FL</sub> gp120 envelope glycoprotein comprising a V3 loop deletion and a C2 domain<sub>(thr260->X)</sub> point mutation, wherein X is an amino acid residue other than threonine. In the preferred embodiment, X is an arginine residue.

- 30 The subject invention additionally provides a recombinant nucleic acid molecule which encodes a mutant HIV-1 gp120 envelope glycoprotein comprising (a) a V3 loop deletion, or (b) a one of the C2, C3 or C4 domain point mutations

discussed supra.

The point mutations in the recombinant nucleic acid molecules described supra are selected based on their  
5 ability to reduce the affinity of the mutant gp120 glycoprotein encoded thereby for CD4. As used herein, the term "reduce the affinity" means to reduce the affinity by at least two-fold.

10 One skilled in the art would know how to make recombinant nucleic acid molecules which encode mutant HIV-1 gp120 envelope glycoproteins comprising a V3 loop deletion and the specific C2, C3 or C4 domain point mutations corresponding to those mutations exemplified in the HIV-1<sub>JR-FL</sub> and HIV-1<sub>LAI</sub>  
15 strains, supra. Furthermore, one skilled in the art would know how to use these recombinant nucleic acid molecules to obtain the proteins encoded thereby, and practice the therapeutic and prophylactic methods of using same, as described herein for the recombinant nucleic acid molecule  
20 which encodes a mutant HIV-1 gp120 envelope glycoprotein comprising a V3 loop deletion and a C4 domain<sub>(W->X)</sub> point mutation.

The subject invention also provides the mutant HIV-1 gp120  
25 envelope glycoprotein encoded by the recombinant nucleic acid molecule of the subject invention.

In accordance with the invention, numerous vector systems for expression of the mutant HIV-1 gp120 envelope  
30 glycoprotein may be employed. For example, one class of vectors utilizes DNA elements which are derived from animal viruses such as bovine papilloma virus, polyoma virus, adenovirus, vaccinia virus, baculovirus, retroviruses (RSV, MMTV or MoMLV), Semliki Forest virus or SV40 virus.

Additionally, cells which have stably integrated the DNA into their chromosomes may be selected by introducing one or more markers which allow for the selection of transfected host cells. The marker may provide, for example, prototrophy to an auxotrophic host, biocide resistance, (e.g., antibiotics) or resistance to heavy metals such as copper or the like. The selectable marker gene can be either directly linked to the DNA sequences to be expressed, or introduced into the same cell by cotransformation. Additional elements may also be needed for optimal synthesis of mRNA. These elements may include splice signals, as well as transcriptional promoters, enhancers, and termination signals. The cDNA expression vectors incorporating such elements include those described by Okayama (22).

The vectors used in the subject invention are designed to express high levels of mutant HIV-1 gp120 envelope glycoproteins in cultured eukaryotic cells as well as efficiently secrete these proteins into the culture medium. The targeting of the mutant HIV-1 gp120 envelope glycoproteins into the culture medium is accomplished by fusing in-frame to the mature N-terminus of the mutant HIV-1 gp120 envelope glycoprotein the tissue plasminogen activator (tPA) prepro-signal sequence.

The mutant HIV-1 gp120 envelope glycoprotein may be produced by a) transfecting a mammalian cell with an expression vector for producing mutant HIV-1 gp120 envelope glycoprotein; b) culturing the resulting transfected mammalian cell under conditions such that mutant HIV-1 gp120 envelope glycoprotein is produced; and c) recovering the mutant HIV-1 gp120 envelope glycoprotein so produced.

Once the expression vector or DNA sequence containing the constructs has been prepared for expression, the expression

vectors may be transfected or introduced into an appropriate mammalian cell host. Various techniques may be employed to achieve this, such as, for example, protoplast fusion, calcium phosphate precipitation, electroporation or other  
5 conventional techniques. In the case of protoplast fusion, the cells are grown in media and screened for the appropriate activity. Expression of the gene encoding a mutant HIV-1 gp120 envelope glycoprotein results in production of the mutant glycoprotein.

10 Methods and conditions for culturing the resulting transfected cells and for recovering the mutant HIV-1 gp120 envelope glycoprotein so produced are well known to those skilled in the art, and may be varied or optimized depending  
15 upon the specific expression vector and mammalian host cell employed.

In accordance with the claimed invention, the preferred host cells for expressing the mutant HIV-1 gp120 envelope  
20 glycoprotein of this invention are mammalian cell lines. Mammalian cell lines include, for example, monkey kidney CV1 line transformed by SV40 (COS-7); human embryonic kidney line 293; baby hamster kidney cells (BHK); Chinese hamster ovary-cells-DHFR (CHO); Chinese hamster ovary-cells DHFR  
25 (DXB11); monkey kidney cells (CV1); African green monkey kidney cells (VERO-76); human cervical carcinoma cells (HELA); canine kidney cells (MDCK); human lung cells (W138); human liver cells (Hep G2); mouse mammary tumor (MMT 060562); mouse cell line (C127); and myeloma cell lines.

30 Other eukaryotic expression systems utilizing non-mammalian vector/cell line combinations can be used to produce the mutant HIV-1 gp120 envelope glycoproteins. These include, but are not limited to, baculovirus vector/insect cell

expression systems and yeast shuttle vector/yeast cell expression systems.

5 Methods and conditions for purifying mutant HIV-1 gp120 envelope glycoproteins from the culture media are provided in the invention, but it should be recognized that these procedures can be varied or optimized as is well known to those skilled in the art.

10 The subject invention further provides a vaccine which comprises a therapeutically effective amount of the mutant HIV-1 gp120 envelope glycoprotein of the subject invention, and an adjuvant.

15 A therapeutically effective amount of the mutant HIV-1 gp120 envelope glycoprotein may be determined according to methods well known to those skilled in the art.

As used herein, adjuvants include, but are not limited to,  
20 alum, Freund's incomplete adjuvant (FIA), Saponin, Quil A, Monophosphoryl lipid A (MPL), and nonionic block copolymers (SAF) such as L-121 (Pluronic; Syntex SAF). In the preferred embodiment, the adjuvant is alum, especially in the form of a thixotropic, viscous, and homogeneous aluminum hydroxide  
25 gel. The vaccine of the subject invention may be administered as an oil in water emulsion. Methods of combining adjuvants with antigens are well known to those skilled in the art.

30 The subject invention further provides a method of treating an HIV-1-infected subject, which comprises immunizing the HIV-1-infected subject with the vaccine of the subject invention, thereby treating the HIV-1-infected subject.

35 As used herein, treating an HIV-1-infected subject with the

vaccine of the subject invention means reducing in the subject either the population of HIV-1 or HIV-1-infected cells, or ameliorating the progression of an HIV-1-related disorder in the subject.

5

As used herein, an "HIV-infected subject" means an individual having at least one of his own cells invaded by HIV-1.

10 As used herein, "immunizing" means administering a primary dose of the vaccine to a subject, followed after a suitable period of time by one or more subsequent administrations of the vaccine, so as to generate in the subject an immune response against the CD4-binding region of the mutant HIV-1  
15 gp120 envelope glycoprotein in the vaccine. A suitable period of time between administrations of the vaccine may readily be determined by one skilled in the art, and is usually in the order of several weeks to months.

20 In the preferred embodiment, the dose of vaccine administered is an amount sufficient to deliver to the subject between 10ug and 1mg of the mutant HIV-1 gp120 envelope glycoprotein.

25 The subject invention further provides a vaccine which comprises a prophylactically effective amount of the mutant HIV-1 gp120 envelope glycoprotein of the subject invention, and an adjuvant.

30 A prophylactically effective amount of the mutant HIV-1 gp120 envelope glycoprotein may be determined according to methods well known to those skilled in the art.

The subject invention further provides a method of reducing  
35 the likelihood of an HIV-1-exposed subject's becoming

infected with HIV-1, which comprises immunizing the HIV-1-exposed subject with the vaccine of the subject invention, thereby reducing the likelihood of the HIV-1-exposed subject's becoming infected with HIV-1.

5

As used herein, the subject's becoming infected with HIV-1 means the invasion of the subject's own cells by HIV-1.

As used herein, reducing the likelihood of a subject's becoming infected with HIV-1 means reducing the likelihood of the subject's becoming infected with HIV-1 by at least two-fold. For example, if a subject has a 1% chance of becoming infected with HIV-1, a two-fold reduction in the likelihood of the subject's becoming infected with HIV-1 would result in the subject's having a 0.5% chance of becoming infected with HIV-1. In the preferred embodiment of this invention, reducing the likelihood of the subject's becoming infected with HIV-1 means reducing the likelihood of the subject's becoming infected with HIV-1 by at least ten-fold.

10  
15  
20

As used herein, an HIV-1-exposed subject is a subject who has HIV-1 present in his body, but has not yet become HIV-1-infected.

25

The subject invention further provides a method of reducing the likelihood of a non-HIV-1-exposed subject's becoming infected with HIV-1, which comprises immunizing the non-HIV-1-exposed subject with the vaccine of the subject invention, thereby reducing the likelihood of the non-HIV-1-exposed subject's becoming infected with HIV-1.

30

As used herein, a non-HIV-1-exposed subject is a subject who does not have HIV-1 present in his body.

35



The subject invention further provides a method of obtaining partially purified antibodies which specifically bind to the CD4-binding domain of HIV-1 gp120 envelope glycoprotein, which method comprises (a) immunizing a non-HIV-1-exposed  
5 subject with the vaccine of the subject invention, (b) recovering from the immunized subject serum comprising said antibodies, and (c) partially purifying said antibodies, thereby obtaining partially purified antibodies which specifically bind to the CD4-binding domain of HIV-1 gp120  
10 envelope glycoprotein. In the preferred embodiment, the subject is a human.

As used herein, partially purified antibodies means a composition which comprises antibodies which specifically  
15 bind to the CD4-binding domain of HIV-1 gp120 envelope glycoprotein, and consists of fewer protein impurities than does the serum from which the anti-CD4-binding domain antibodies are derived. A protein impurity means a protein other than the anti-CD4-binding domain antibodies. For  
20 example, the partially purified antibodies might be an IgG preparation.

Methods of recovering serum from a subject are well known to those skilled in the art. Methods of partially purifying  
25 antibodies are also well known to those skilled in the art, and include, by way of example, filtration, ion exchange chromatography, and precipitation.

In one embodiment, the partially purified antibodies  
30 comprise an immune globulin (IG) preparation. IG can be purified from serum by a two-step process. Initially, serum is fractionated by the cold ethanol method of Cohn, et al. (29). Cohn Fraction II has as its main protein component IgG immunoglobulin present as monomers, dimers and  
35 aggregates. Fraction II is then purified to produce IVIG

(immune globulin intravenous) using a variety of purification methods which include, for example, ion exchange, DEAE chromatography, acid pH 4.25 diafiltration, PEG precipitation or Pepsin treatment. The final product is  
5 stabilized (e.g., glucose + NaCl) and the final IgG concentration is fixed at between about 3% and about 6%.

The subject invention further provides the partially purified antibodies produced by the method of the subject  
10 invention.

The subject invention further provides a pharmaceutical composition, which comprises a therapeutically effective amount of the partially purified antibodies of the subject  
15 invention, and a pharmaceutically acceptable carrier.

A therapeutically effective amount of the partially purified antibodies of the subject invention may be determined according to methods well known to those skilled in the art.  
20

Pharmaceutically acceptable carriers are well known to those skilled in the art and include, but are not limited to, 0.01-0.1M and preferably 0.05M phosphate buffer or 0.8% saline. Additionally, such pharmaceutically acceptable  
25 carriers may be aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include  
30 water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's or fixed oils. Intravenous vehicles include fluid  
35 and nutrient replenishers, electrolyte replenishers such as

(immune globulin intravenous) using a variety of purification methods which include, for example, ion exchange, DEAE chromatography, acid pH 4.25 diafiltration, PEG precipitation or Pepsin treatment. The final product is  
5 stabilized (e.g., glucose + NaCl) and the final IgG concentration is fixed at between about 3% and about 6%.

The subject invention further provides the partially purified antibodies produced by the method of the subject  
10 invention.

The subject invention further provides a pharmaceutical composition, which comprises a therapeutically effective amount of the partially purified antibodies of the subject  
15 invention, and a pharmaceutically acceptable carrier.

A therapeutically effective amount of the partially purified antibodies of the subject invention may be determined according to methods well known to those skilled in the art.

20

Pharmaceutically acceptable carriers are well known to those skilled in the art and include, but are not limited to, 0.01-0.1M and preferably 0.05M phosphate buffer or 0.8% saline. Additionally, such pharmaceutically acceptable  
25 carriers may be aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include  
30 water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's or fixed oils. Intravenous vehicles include fluid  
35 and nutrient replenishers, electrolyte replenishers such as

those based on Ringer's dextrose, and the like. Preservatives and other additives may also be present, such as, for example, antimicrobials, antioxidants, chelating agents, inert gases and the like.

5

The subject invention further provides a method of treating an HIV-1-infected subject, which comprises administering to the subject a dose of the pharmaceutical composition of the subject invention effective to reduce the population of HIV-1-infected cells in the HIV-1-infected subject, thereby treating the HIV-1-infected subject.

As used herein, administering may be effected or performed using any of the various methods known to those skilled in the art. The administering may comprise administering intravenously. The administering may also comprise administering intramuscularly. The administering may further comprise administering subcutaneously.

The dose of the pharmaceutical composition of the subject invention effective to reduce the population of HIV-1-infected cells in the HIV-1-infected subject may be readily determined using methods well known to those skilled in the art. In the preferred embodiment, the dose is sufficient to deliver to the subject between about 10 mg/kg and 150mg/kg of protein if administered intramuscularly. In the preferred embodiment, the dose is sufficient to deliver to the subject between about 100 mg/kg and 2g/kg of protein if administered intravenously.

30

The subject invention further provides a method of treating an HIV-1-infected subject, which comprises administering to the subject a dose of the pharmaceutical composition of the subject invention effective to reduce the population of HIV-1 in the HIV-1-infected subject, thereby treating the HIV-1-

35

those based on Ringer's dextrose, and the like. Preservatives and other additives may also be present, such as, for example, antimicrobials, antioxidants, chelating agents, inert gases and the like.

5

The subject invention further provides a method of treating an HIV-1-infected subject, which comprises administering to the subject a dose of the pharmaceutical composition of the subject invention effective to reduce the population of HIV-1-infected cells in the HIV-1-infected subject, thereby treating the HIV-1-infected subject.

As used herein, administering may be effected or performed using any of the various methods known to those skilled in the art. The administering may comprise administering intravenously. The administering may also comprise administering intramuscularly. The administering may further comprise administering subcutaneously.

The dose of the pharmaceutical composition of the subject invention effective to reduce the population of HIV-1-infected cells in the HIV-1-infected subject may be readily determined using methods well known to those skilled in the art. In the preferred embodiment, the dose is sufficient to deliver to the subject between about 10 mg/kg and 150mg/kg of protein if administered intramuscularly. In the preferred embodiment, the dose is sufficient to deliver to the subject between about 100 mg/kg and 2g/kg of protein if administered intravenously.

30

The subject invention further provides a method of treating an HIV-1-infected subject, which comprises administering to the subject a dose of the pharmaceutical composition of the subject invention effective to reduce the population of HIV-1 in the HIV-1-infected subject, thereby treating the HIV-1-

35

infected subject.

The dose of the pharmaceutical composition of the subject invention effective to reduce the population of HIV-1 in the  
5 HIV-1-infected subject may be readily determined using methods well known to those skilled in the art. In the preferred embodiment, the dose is sufficient to deliver to the subject between about 10 mg/kg and 150mg/kg of protein if administered intramuscularly. In the preferred  
10 embodiment, the dose is sufficient to deliver to the subject between about 100 mg/kg and 2g/kg of protein if administered intravenously.

The subject invention further provides a composition which  
15 comprises a prophylactically effective amount of the partially purified antibodies of the subject invention, and a pharmaceutically acceptable carrier.

A prophylactically effective amount of the partially  
20 purified antibodies of the subject invention may be determined according to methods well known to those skilled in the art.

The subject invention further provides a method of reducing  
25 the likelihood of an HIV-1-exposed subject's becoming infected with HIV-1, which comprises administering to the HIV-1-exposed subject a dose of the composition of the subject invention effective to reduce the population of HIV-1 in the HIV-1-exposed subject, thereby reducing the  
30 likelihood of the subject's becoming infected with HIV-1.

In one embodiment, the subject is a medical practitioner. The medical practitioner may be a medical practitioner exposed to an HIV-1-containing bodily fluid. As used herein,  
35 the term "medical practitioner" includes, but is in no way

limited to, doctors, dentists, surgeons, nurses, medical laboratory assistants, and students in health care programs.

In another embodiment, the subject is a newborn infant. The  
5 newborn infant may be a newborn infant born to an HIV-1-infected mother.

The dose of the composition of the subject invention effective to reduce the population of HIV-1 in the HIV-1-  
10 exposed subject may be readily determined using methods well known to those skilled in the art. In the preferred embodiment, the dose is sufficient to deliver to the subject between about 10mg/kg and 150mg/kg of protein if administered intramuscularly. In the preferred embodiment,  
15 the dose is sufficient to deliver to the subject between about 100 mg/kg and 2g/kg of protein if administered intravenously.

The vaccines and pharmaceutical compositions of the subject  
20 invention may also ameliorate the progression of an HIV-1-related disorder in a subject to whom the vaccines or pharmaceutical compositions were administered while the subject was either non-HIV-1-exposed or HIV-1-exposed, but not yet HIV-1-infected.

25 Finally, the subject invention provides a method of reducing the likelihood of a non-HIV-1-exposed subject's becoming infected with HIV-1 as a result of exposure thereto during an incident wherein there is an increased risk of exposure  
30 to HIV-1, which comprises administering to the subject immediately prior to the incident a dose of the composition of the subject invention effective to reduce the population of HIV-1 to which the subject is exposed during the incident, thereby reducing the likelihood of the subject's  
35 becoming infected with HIV-1. In one embodiment, the

subject is a medical practitioner.

An incident wherein there is an increased risk of exposure to HIV-1 includes, for example, receiving a blood  
5 transfusion, sexual contact with an HIV-1-infected individual, and performing a HIV-1-containing bodily fluid-exposing medical procedure.

As used herein, "immediately prior to the incident" means  
10 within one month of the incident. In the preferred embodiment, "immediately prior to the incident" means within one day of the incident.

The dose of the composition of the subject invention  
15 effective to reduce the population of HIV-1 to which the subject is exposed during the incident may be readily determined using methods well known to those skilled in the art. In the preferred embodiment, the dose is sufficient to deliver to the subject between about 10mg/kg and 150mg/kg of  
20 protein if administered intramuscularly. In the preferred embodiment, the dose is sufficient to deliver to the subject between about 100mg/kg and 2g/kg of protein if administered intravenously.

25 One embodiment of this invention is a method of substantially reducing the likelihood of a non-infected medical practitioner's becoming infected with HIV-1 during a bodily fluid-exposing medical procedure involving a patient, which comprises administering to the patient during  
30 a suitable time period an amount of the composition of the subject invention effective to substantially reduce the likelihood of the non-infected medical practitioner's becoming infected with HIV-1 by virtue of contact with the patient's bodily fluid during the medical procedure.



As used herein, a bodily fluid is any fluid which is present in the human body and is capable of containing infectious HIV-1 in an HIV-1-infected patient. Bodily fluids include, but are not limited to, saliva, cerebrospinal fluid, tears, vaginal secretions, urine, alveolar fluid, synovial fluid and pleural fluid.

Another embodiment of this invention is a method of substantially reducing the likelihood of a non-HIV-1-infected newborn infant's becoming infected with HIV-1 prior to or during birth from an HIV-1-infected mother, which comprises administering to the mother prior to birth an amount of the composition of the subject invention effective to substantially reduce the likelihood of the non-HIV-1-infected newborn infant's becoming infected with HIV-1 by virtue of contact with the patient's bodily fluid.

In order to facilitate an understanding of the Experimental Details section which follows, certain frequently occurring methods and/or terms are best described in Maniatis et al. (23).

This invention will be better understood by reference to the Experimental Details which follow, but those skilled in the art will readily appreciate that the specific experiments detailed are only illustrative of the invention as described more fully in the claims which follow thereafter.

## Experimental Details

### Nomenclature

As used herein, V3<sup>(-)</sup> indicates a V3 loop deletion from HIV-1 gp120 envelope glycoprotein. As used herein, CD4<sup>(-)</sup> indicates a point mutation in the C4 domain of HIV-1 gp120 envelope glycoprotein which mutation inhibits CD4 binding to the mutant HIV-1 gp120 envelope glycoprotein. The structure of HIV-1 gp120 envelope glycoprotein is shown in Figure 1.

### Materials and Methods

#### 1. Construction of PPI4-tPA-gp120<sub>LAI</sub> expression vector.

An expression vector was constructed that consisted of the cytomegalovirus major immediate early (CMV MIE) promoter/enhancer linked to the HIV-1<sub>LAI</sub> env gene, which gene had its signal sequence replaced by the tPA signal sequence. The CMV MIE promoter/enhancer sequences were derived from pSVCC1 (24) consisting of 1580 base pairs of contiguous DNA that is immediately 5' to the initiator ATG. In sequential order, the functional domains of the CMV promoter are: the promoter/enhancer region; a transcriptional initiator site; exon A (a non-coding exon); intron A; and 17 nucleotides of exon B (non-coding sequences). The viral promoter sequences were ligated to a gene construct consisting of the nucleotide sequences encoding amino acids -35 to -1 of human tPA (25) fused in-frame to HIV-1<sub>LAI</sub> env amino acids 31 through 515, ending with a TGA stop codon. The construction was performed in two parts. The majority of the CMV promoter could be isolated as a 1560 bp Hinc II/Pst I fragment which was ligated to a Pst I/Not I 1590 bp DNA fragment that contained the remainder of the CMV promoter, the initiator ATG, the tPA signal sequence and the mature HIV-1<sub>LAI</sub> env protein coding sequence.

The latter fragment was assembled using the polymerase chain reaction as follows. Primer 1 (GATCCTGCAGTCACCGTCCTTGACACGATGGATGCAATGAAGAGA) and primer 2 (AAGTCTTCTCCTCGGTCTTGTCTTTTAAACACCCAG) were used to amplify the nucleic acid sequences encoding the tPA signal sequence amino acids -35 to -1 from plasmid pMAM neo-s (Clonetech), thus producing a 150 bp fragment. A second 1440 bp DNA fragment was amplified using primer 3 (TTCAGAAGAGGAGCCAGAACAGAAAAATTGTGGGTC), primer 4 (GGAAAAAAGCGGCCGCTCATTTTTCTCTCTGCACCACTC), and pENV (26) as a template. The PCR fragments were pooled, desalted, and excess primer removed by ultrafiltration through a centricon-100 unit (Amicon). An aliquot of the pooled material was then subjected to a second round of amplification in the presence of primers 1 and 4 to produce a 1590 bp fragment, which was then digested with Pst I and Not I. The CMV promoter fragment and the HIV-1<sub>LAI</sub> env fragment were then ligated together, and the entire transcription unit subcloned into PPI4, which is a eukaryotic shuttle vector that contains an ampicillin resistance gene, an SV40 origin of replication and a DHFR gene whose transcription is driven by the  $\beta$ -globin promoter. The final construct, PPI4-tPA-gp120<sub>LAI</sub>, is shown in Figure 2.

The expression vector is then used as the prototype vector for the expression of gp120 proteins that are derived from other HIV-1 strains or mutated as described in the methods section. The vector was constructed so that unique Nar I and Not I sites flank the gp120 sequence, thus facilitating the removal of the gp120 gene cassette and the subsequent insertion of other gene cassettes (Figure 2).

## 2. Expression of HIV-1<sub>LAI</sub> gp120 in mammalian cells.

### a. Transient expression.

CosM5 cells grown in DMEM containing 10% fetal calf serum

were split to 75% confluence. On the following day, the cells were transfected for 16-20 hours with 10 micrograms of CsCl-purified PPI4-tPA-gp120<sub>LAI</sub> DNA by the standard CaPO<sub>4</sub> (5) precipitation technique. After transfection, fresh medium was added to the cells. Analysis of the products synthesized 96-120 hours post-transfection was performed by radiolabelling the transfectants with <sup>35</sup>S-cysteine for 12-18 hours, followed by precipitation of media using a CD4-immunoglobulin-Protein A-Sepharose complex, followed by SDS-PAGE under reducing conditions (Figure 4).

b. Stable expression.

Dhfr<sup>r</sup> Chinese hamster ovary cells (CHO) were transfected with 20 micrograms of CsCl-purified DNA. Approximately 3-5 days post-transfection, cells were placed in selective medium (nucleoside-free alpha MEM containing 10% dialyzed fetal calf serum). Approximately 10-15 days post-selection, individual cell clones were picked. Media was analyzed for gp120 expression by radiolabelling the cells with <sup>35</sup>S-cysteine for 12-18 hours, followed by precipitation of media using a CD4-immunoglobulin-Protein A-Sepharose complex, followed in turn by SDS-PAGE under reducing conditions (Figure 6). The levels of gp120 in the media of these clones were also quantitated (Figure 5) by ELISA performed as follows. The method involves coating 96-well plates overnight with sheep polyclonal IgG against the highly conserved C-terminus of gp120 (D7234, Aalto Bioreagents). After washing, dilutions of a standard gp120 preparation in cell growth medium, or supernatant from the stably-transfected cells, were incubated for 1 hour. The plates were washed again, and incubated for one hour with a horseradish peroxidase-conjugated anti-gp120 monoclonal antibody (9204, DuPont). Following a final wash, the peroxidase substrate OPD (DuPont) was added and the amount

of gp120 determined by comparing absorbance of unknowns with a standard curve. Standards were prepared from purified gp120 made in CHO cells, a small quantity of which was obtained from Celltech Ltd. Clones expressing the highest  
5 levels were subjected to successive rounds of amplification of the newly introduced DNA sequences in increasing concentrations of methotrexate. Stable CHO cell lines were thus generated which secrete at least 1 microgram/milliliter of HIV-1<sub>LAI</sub> gp120.

10

3. Construction of PPI4-tPA-gp120<sub>JR-FL</sub>

a. The HIV-1<sub>LAI</sub> gp120 env nucleotide sequence in PPI4-tPA-gp120<sub>LAI</sub> was replaced by the nucleotide sequence encoding the mature gp120<sub>JR-FL</sub> protein. Using the polymerase chain  
15 reaction, the JR-FL sequences were amplified from pUC112-1 (27) using primer 5 (GATCGGCGCCAGAGTAGAAAAGTTGTGGGTCAC) and primer 4. The PCR fragment was digested with the restriction endonucleases Nar I and Not I, and the fragment subcloned in between the Nar I and Not I sites in PPI4-tPA-  
20 gp120<sub>LAI</sub> to generate PPI4-tPA-gp120<sub>JR-FL</sub> (Figure 7).

b. Transient expression.

CosM5 cells grown in DMEM containing 10% fetal calf serum were split to 75% confluence. On the following day, the  
25 cells were transfected for 16-20 hours with 10 micrograms of CsCl-purified PPI4-tPA-gp120<sub>JR-FL</sub> DNA by the standard CaPO<sub>4</sub> (5) precipitation technique. After transfection, fresh medium was added to the cells. Analysis of the products synthesized 96-120 hours post-transfection was performed by  
30 radiolabelling the transfectants with <sup>35</sup>S-cysteine for 12-18 hours, followed by precipitation of media using a CD4-immunoglobulin-Protein A-Sepharose complex, followed by SDS-PAGE under reducing conditions (Figure 4).

#### 4. Construction of PPI4-tPA-gp120<sub>LAI</sub>-V3<sup>(c)</sup>.

The V3 loop in tPA-gp120<sub>LAI</sub> consists of amino acids Cys<sub>306</sub> through Cys<sub>333</sub>. In the V3<sup>(c)</sup> mutant, the amino acids in between these cysteines are replaced by the pentapeptide sequence Thr-Gly-Ala-Gly-His. Using the Transformer Site-Directed Mutagenesis Kit (Clontech), the V3 loop sequence in PPI4-tPA-gp120<sub>LAI</sub> is altered using the mutagenic primer 6 (CTGTAGAAATTAATTGTACAGGTGCTGGACATTGTAACATTAGTAGAGC) and primer 7 (CTCGAGCATGCATTCGAAGCTCGCTGATC) as a selection primer. Primer 7 changes a unique Xba I site in the backbone of the parent PPI4 plasmid into a unique BstB I site. Briefly, the mutagenesis method requires incubating of the parent plasmid with the mutagenic primer and the selection primer, denaturing at 100°C for 3 minutes and then chilling on ice. In the presence of buffered deoxynucleotide triphosphates and T4 DNA polymerase, the primers are allowed to initiate the polymerization of one strand of plasmid DNA. T4 DNA ligase is used to seal the newly synthesized DNA strand to form a covalently closed circle. Hybrid plasmids are then transformed into a MutS strain of *E. coli* that is deficient in mismatch repair. After allowing for the growth of transformed cells, DNA is purified from the cells and digested with the selection restriction endonuclease, in this case Xba I. Parental plasmids are cleaved by Xba I while the mutant plasmid remains resistant to cleavage by virtue of the Xba I to BstB I conversion. Digested DNA is then used to transform *E. coli*, and colonies harboring the mutant plasmid are picked. Multiple mutagenic primers can be used in a single round of mutagenesis. The amino acid sequence of the modified protein is shown in Figure 8.

#### 5. Construction of PPI4-tPA-gp120<sub>IR-FL</sub>-V3<sup>(c)</sup>.

The V3 loop in tPA-gp120<sub>IR-FL</sub> consists of amino acids Cys<sub>293</sub>

through Cys<sub>37</sub>. In the V3<sup>(c)</sup> mutant, the amino acids in between these cysteines are replaced by the pentapeptide sequence Thr-Gly-Ala-Gly-His. Using the Transformer Site-Directed Mutagenesis Kit (Clonetech), the V3 loop sequence in PPI4-tPA-gp120<sub>JR-FL</sub> is altered using the mutagenic primer 6 (CTGTAGAAATTAATTGTACAGGTGCTGGACATTGTAACATTAGTAGAGC) and primer 7 as a selection primer. The amino acid sequence of the modified protein is shown in Figure 9.

10 6. Construction of PPI4-tPA-gp120<sub>LAI</sub>-CD4<sup>(c)</sup>.

Using the Transformer Site-Directed Mutagenesis Kit (Clonetech), the selection primer 7, and the mutagenic primer 8 (CAATTTATAAACATGGTGCAGGAAGTAGG), Trp<sub>437</sub> of tPA-gp120<sub>LAI</sub>, which is in an equivalent position to the tryptophan residue in the HXBc2 strain of HIV-1, is mutated to a Val in the expression vector PPI4-tPA-gp120<sub>LAI</sub> to generate PPI4-tPA-gp120<sub>LAI</sub>-CD4<sup>(c)</sup>. The sequence for gp120<sub>LAI</sub>-CD4<sup>(c)</sup> is shown in Figure 12.

20 7. Construction of PPI4-tPA-gp120<sub>JR-FL</sub>-CD4<sup>(c)</sup>.

In a fashion similar to that described above, Trp<sub>424</sub> of tPA-gp120<sub>JR-FL</sub> is mutated to a Val in the expression vector PPI4-tPA-gp120<sub>JR-FL</sub> using the selection primer 7 and the mutagenic primer 9 (CAAATTATAAACATGGTGCAGGAAGTAGG) to generate PPI4-tPA-gp120<sub>JR-FL</sub>-CD4<sup>(c)</sup>. The sequence for gp120<sub>JR-FL</sub>-CD4<sup>(c)</sup> is shown in Figure 13.

8. Construction of PPI4-tPA-gp120<sub>LAI</sub>-V3<sup>(c)</sup>-CD4<sup>(c)</sup>.

The tPA-gp120<sub>LAI</sub> double mutant, V3<sup>(c)</sup>-CD4<sup>(c)</sup>, is constructed by including the mutagenic primers 6 and 8, and the selection primer 7 simultaneously in the reaction tube with PPI4-tPA-gp120<sub>LAI</sub> as the DNA template. The final construct is named PPI4-tPA-gp120<sub>LAI</sub>-V3<sup>(c)</sup>-CD4<sup>(c)</sup>, and its sequence is shown in figure 10.

9. Construction of PPI4-tPA-gp120<sub>JR-FL</sub>-V3<sup>(-)</sup>-CD4<sup>(-)</sup>.

The tPA-gp120<sub>JR-FL</sub> double mutant, V3<sup>(-)</sup>-CD4<sup>(-)</sup>, is constructed by including the mutagenic primers 6 and 9, and the selection primer 7 simultaneously in the reaction tube with PPI4-tPA-gp120<sub>JR-FL</sub> as the DNA template. The final construct is named PPI4-tPA-gp120<sub>JR-FL</sub>-V3<sup>(-)</sup>-CD4<sup>(-)</sup>, and its sequence is shown in figure 11.

10. Expression of mutant HIV-1 gp120 in mammalian cells.

10 a. Transient expression.

CosM5 cells grown in DMEM containing 10% fetal calf serum are split to 75% confluence. On the next day, the cells are transfected for 16-20 hours with 10 micrograms of CsCl-purified mutant HIV-1 DNA by the standard CaPO<sub>4</sub> (5) precipitation technique. After transfection, fresh medium is added to the cells. Analysis of the products synthesized 96-120 hours post-transfection is performed by radiolabelling the transfectants with <sup>35</sup>S-cysteine for 12-18 hours, followed by precipitation of media using a sheep polyclonal IgG against the highly conserved C-terminus of gp120.

b. Stable expression.

Dhfr<sup>r</sup> Chinese hamster ovary cells (CHO) are transfected with 20 micrograms of CsCl-purified DNA encoding the native or mutant HIV-1 gp120 glycoproteins. Approximately 3-5 days post-transfection, cells are placed in selective medium (nucleoside-free alpha MEM containing 10% dialyzed fetal calf serum). Approximately 10-15 days post-selection, individual cell clones are picked. Media is analyzed for gp120 expression by radiolabelling the cells with <sup>35</sup>S-cysteine for 12-18 hours, followed by quantitative immunoprecipitation of media using a sheep polyclonal IgG against the highly conserved C-terminus of gp120, followed



in turn by SDS-PAGE under reducing conditions. Alternatively, one can quantitate the level of gp120 by ELISA performed as follows. The method involves coating 96-well plates overnight with sheep polyclonal IgG against the highly conserved C-terminus of gp120 (D7234, Aalto Bioreagents). After washing, dilutions of a standard gp120 preparation in cell growth medium, or supernatant from the stably-transfected cells, are incubated for 1 hour. The plates are washed again, and incubated for one hour with a human MoAb (F105, AIDS Research & Reference Reagent Program, No. 857). The plates are washed again, and incubated again for 1 hour with a horseradish-peroxidase-conjugated goat anti-human IgG (Cappel). Following a final wash, the peroxidase substrate OPD (DuPont) is added and the amount of gp120 determined by comparing absorbance of unknowns with a standard curve. Standards are prepared from purified gp120 made in CHO cells, a small quantity of which is obtained from Celltech Ltd. Clones expressing the highest levels are subjected to successive rounds of amplification of the newly introduced DNA sequences in increasing concentrations of methotrexate. Stable CHO cell lines are thus generated which secrete at least 1 microgram/milliliter of mutant HIV-1 gp120.

25 11. Purification of HIV-1 gp120 proteins.

A one-step immunoaffinity procedure is used to purify the recombinant gp120 molecules described. Briefly, culture supernatant is collected and clarified by centrifugation. An immunoaffinity column consisting of a matrix coupled to a sheep polyclonal anti-gp120 IgG (D7234, Aalto Bioreagents) directed against the highly conserved C-terminal end (APTKAKRRVVQREKR) of gp120 is used to specifically adsorb gp120 from the cell culture media. This antisera recognizes native gp120, the V3 loop deletion mutants, and the CD4<sup>(-)</sup>

mutants since the C-terminal ends of these molecules remain unaltered. The bound gp120 is then eluted with 2M MgCl<sub>2</sub>, concentrated by Amicon filtration, and dialyzed into 10 mM HEPES, pH 7.0. The purity of the proteins is determined by  
5 SDS-PAGE and silver staining.

12. Characterization of recombinant HIV-1 gp120 proteins.  
The purified glycoproteins are subjected to extensive biochemical and immunologic characterization. The integrity  
10 of the proteins is monitored by SDS-PAGE and silver staining under reducing and non-reducing conditions. The glycoproteins are deglycosylated by treatment with the enzyme N-glycosidase F which cleaves N-linked oligo-saccharides, and are assayed by SDS-PAGE and silver staining  
15 to monitor molecular weight shifts. The purified glycoproteins are also tested for reactivity with several well characterized anti-gp120 monoclonal antibodies that recognize both linear and discontinuous epitopes. The binding affinity to sCD4 is estimated using an ELISA assay.

20 The purified proteins HIV-1 gp120<sub>LAI</sub>, gp120<sub>LAI</sub>-V3<sup>(-)</sup>, gp120<sub>LAI</sub>-V3<sup>(-)</sup>-CD4<sup>(-)</sup>, gp120<sub>JR-FL</sub>, gp120<sub>JR-FL</sub>-V3<sup>(-)</sup>, and gp120<sub>JR-FL</sub>-V3<sup>(-)</sup>-CD4<sup>(-)</sup>, were tested for their ability to bind cell surface human CD4. DG44 #3 cells, a recombinant cell line designed to express  
25 human CD4 on the membrane surface, were grown in T flasks and trypsinized. 5 X 10<sup>5</sup> cells/experiment were aliquoted into FACS buffer (PBS + 2% BSA and 0.1% NaN<sub>3</sub>), washed several times in the same buffer, and then incubated with 100 ul of a solution of purified gp120 protein at 5ug/ml in  
30 FACS buffer at 37°C for 2 hr. The cells were washed in FACS buffer, and then incubated in 100 ul solution containing 5ug/ml sheep polyclonal IgG against the highly conserved C-terminus of gp120 in FACS buffer at 37°C for 2 hr. The cells were washed in FACS buffer then incubated in 100 ul

solution containing FITC-labeled rabbit anti-sheep IgG polyclonal antibody at 37°C for 2 hr. The cells were washed with FACS buffer and then resuspended in 500 ul FACS buffer. The cells were then analyzed on a Becton Dickinson FACScan according to the manufacturer's instructions. As a control for expression of CD4 on the DG44 #3 cells, FITC-labeled OKT4A (Becton Dickinson) was used.

13. A protocol for inoculation of animals with the mutant HIV-1 gp120 envelope glycoproteins.

Alum is used as an adjuvant during the inoculation series. The inoculum is prepared by dissolving the mutant HIV-1 gp120 envelope glycoprotein antigen in physiologic saline at a final antigen concentration of 100 ug/ml. Preformed alum (aluminum hydroxide gel) is added to the solution to a final level of 500 ug/ml aluminum. The antigen is allowed to adsorb onto the alum gel for two hours at room temperature. Following adsorption, the gel with the antigen is washed twice with physiologic saline and resuspended in the saline to a protein concentration of 100 ug/ml.

Monkeys and/or Guinea Pigs are individually inoculated with four 100 ug doses of the mutant HIV-1 gp120 envelope glycoprotein antigen adsorbed onto alum. Each dose is injected intramuscularly. The doses are delivered one or five months apart (week 0, 4, 8 and 28). the animals are bled at intervals of two or four weeks. Serum samples are prepared from each bleed to assay for the development of specific antibodies as described in the subsequent sections.

30

14. Analysis of sera for anti-mutant HIV-1 gp120 envelope glycoprotein IgG antibodies.

Each serum sample is analyzed by ELISA. Polystyrene microtiter plates are coated with 0.5 ug per well of pure mutant HIV-1 gp120 envelope glycoprotein in phosphate-

35

buffered physiological saline (PBS) at 4°C. Each well is then washed with PBS containing 0.5% TWEEN-20 (PBS-TW). Test serum, diluted serially in PBS-TW, is added to the mutant HIV-1 gp120 envelope glycoprotein-containing wells and allowed to react with the adsorbed mutant HIV-1 gp120 envelope glycoprotein for one hour at 37°C. The wells are then washed extensively in PBS-TW. Each well then receives 0.1% p-nitrophenyl phosphate in 10% diethanolamine, pH 9.8, containing 0.5 mM  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ . The ensuing reaction is allowed to proceed at room temperature for 30 minutes, at which time it is terminated by the addition of 3.0 N NaOH.

The greater the interaction of antibodies in the test serum with the mutant HIV-1 gp120 envelope glycoprotein, the greater is the amount of alkaline phosphatase bound onto the well. The phosphatase enzyme mediates the breakdown of p-nitrophenyl phosphate into a molecular substance which absorbs light at a wavelength of 405 nm. Hence, there exists a direct relationship between the absorbance at 405 nm of light at the end of the ELISA reaction and the amount of mutant HIV-1 gp120 envelope glycoprotein-bound antibody. All animals inoculated with mutant HIV-1 gp120 envelope glycoprotein whose serum reacts specifically with the mutant HIV-1 gp120 envelope glycoprotein in the ELISA have a positive antibody response against mutant HIV-1 gp120 envelope glycoprotein.

15. Analysis of sera for activity which specifically neutralizes HIV-1 infectivity.

30 Virus-neutralizing activity is determined with an assay based on the use of multiplicity curves in which the ratio of infectious virus surviving antibody treatment ( $V_a$ ) is compared to infectious virus in uninhibited cultures ( $V_o$ ) at various dilutions of antisera. The neutralization titer of

the sera is then interpolated as that sera dilution which yields one log reduction in infectious titer (i.e.,  $V_0/V_1 = 0.1$ ). Briefly, 4-fold dilutions of virus (laboratory-adapted and primary isolates) are prepared to yield  
5 infectious doses of 0.1 to 100 TCID<sub>50</sub> (Tissue Culture Infection Dose) in 20 ul. Serial 3-fold dilutions of sera are also prepared and 20 ul of each serum dilution are incubated with each dilution of virus in duplicate for 60 minutes at room temperature in a 96-well microtiter plate.  
10 20 ul of AA5 cells (PHA stimulated PBMCs for primary HIV-1 isolates) are then added to the serum/virus mixtures. Cells are cultured for 7 days by the addition of fresh medium every other day. On the seventh day, supernatant from each well is removed and tested for the presence of reverse  
15 transcriptase (RT). Infection in each well is then scored as either positive or negative based on the RT counts, and the infectious dose of virus in each treatment group is calculated using the Reed and Muench (28) formula. The neutralization titers represent the reciprocal serum  
20 dilution required to reduced infectious dose of virus by one log. The above culture time is for the prototypic HIV-1<sub>LAI</sub> isolate tested on the AA5 cell line. In the case of primary isolates, the termination date is usually 11-14 days. Culture conditions for PBMCs is not as demanding since  
25 doubling time is restricted. In the case of PBMCs, one day PHA stimulations are used at a final concentration of  $1.5 \times 10^6$ /ml on day 0. Half that number of fresh PBMCs are then added again on days 4 and 8. This multiple addition of PBMCs is meant to amplify virus output upon successful  
30 infection so that the readout RT signal is strong. Again, the final readout titer for the primary isolate/PBMC is the reciprocal serum dilution which reduces infectious titer by one log.

16. Passive hyperimmune therapy.

Non-HIV-1-infected humans are immunized with the mutant HIV-1 gp120 envelope glycoprotein antigens according to a protocol similar to that described above in section 12. For  
5 passive hyperimmune therapy in HIV-1-infected individuals, blood plasma is taken from mutant HIV-1 gp120 envelope glycoprotein immunized, non-HIV-1-infected human donors whose plasma has high levels of neutralizing antibodies. The plasma is pooled from several donors, purified to remove  
10 nonimmunoglobulin proteins and is then sterilized to kill any other viruses or pathogens. The treated plasma is then injected into individuals infected with HIV-1, with repeated injections every week, every two weeks, or every month.

### Results

Eukaryotic expression vectors designed to express high levels of HIV-1<sub>LAI</sub> gp120 and HIV-1<sub>JR-FL</sub> gp120 were constructed.

5 The CMV MIE promoter/enhancer was used to drive the transcription of a gene fusion consisting of the human tPA signal sequence fused to mature gp120 (Figures 2 and 7). The complete sequence of the transcription unit from the Hinc II site of the CMV promoter/enhancer to the Not I site  
10 just 3' from the stop codon in gp120 is shown in figure 3. This vector was used to transfect COSM5 cells in a transient assay. The transfected cells were labeled with <sup>35</sup>S-cysteine and the media immunoprecipitated with a CD4-immunoglobulin-Protein A-Sepharose complex. The precipitated products were  
15 analyzed using a reducing 10% SDS-PAGE gel and autoradiography (Figure 4). A 120 kD band was detected when PPI4-tPA-gp120<sub>LAI</sub> was used to transfect COS cells (lane 3). A band migrating with a slightly lower molecular mass was detected when PPI4-tPA-gp120<sub>JR-FL</sub> was used to transfect COS  
20 cells (lane 4). No radiolabeled products were detected in the mock infected cells. Using a sheep polyclonal antibody directed against the highly conserved C-terminal end of HIV-1 gp120 in an ELISA assay, the level of expression of HIV-1 gp120 was determined to be 2350 ng/ml.

25

The PPI4-tPA-gp120<sub>LAI</sub> vector was then used to stably transfect the dhfr CHO cell line DXB11. Two days post-transfection, the cells were plated at low density in nucleoside-free medium. Eight days post-transfection,  
30 surviving clones were isolated and expanded. Individual primary transfectants were tested for gp120 expression using the ELISA method described in the methods section. Several primary CHO transfectants expressed significant quantities (10-120 ng/ml) of gp120 (Figure 5). Three of the highest

expressing clones were then subjected to increasing concentrations of methotrexate in order to amplify, in tandem, the copy number of the dhfr and gp120 genes. Cell lines were established that express high levels of gp120 with rates of secretion greater than 1 mg/liter. These were then used to purify gp120 to homogeneity.

Six CHO cell lines were established, using the procedures described in the methods sections, that express high levels of the following proteins: HIV-1 gp120<sub>LAI</sub>, gp120<sub>LAI</sub>-V3<sup>(-)</sup>, gp120<sub>LAI</sub>-V3<sup>(-)</sup>-CD4<sup>(-)</sup>, gp120<sub>JR-FL</sub>, gp120<sub>JR-FL</sub>-V3<sup>(-)</sup>, and gp120<sub>JR-FL</sub>-V3<sup>(-)</sup>-CD4<sup>(-)</sup>. Metabolic labeling of these cells with <sup>35</sup>S-cysteine followed by immunoprecipitation with the human monoclonal antibody F105 and analyzed by SDS-PAGE and autoradiography showed the presence of the gp120 proteins in the culture supernatant (Figure 14). From these cell lines the gp120 proteins were purified to homogeneity. Analysis by SDS-PAGE followed by silver-staining showed the purity of these proteins to be greater than 90% (Figure 15).

It was shown by FACScan analysis that the two CD4 binding mutants HIV-1gp120<sub>LAI</sub>-V3<sup>(-)</sup>-CD4<sup>(-)</sup> and HIV-1 gp120<sub>JR-FL</sub>-V3<sup>(-)</sup>-CD4<sup>(-)</sup> had no appreciable binding to recombinant cell lines designed to express high levels of human CD4 on their membrane surface (Figure 16, panel 4 and data not shown, respectively).



### Discussion

The advantage of using the mutant HIV-1 gp120 envelope glycoproteins as immunogens is that these proteins will not elicit an immune response against the V3 loop, a highly immunodominant epitope on gp120. This is significant because the V3 loop may skew the humoral immune response away from discontinuous epitopes in the CD4-binding site. Mutant HIV-1 gp120 envelope glycoproteins having partial and total V3 loop deletions have been made (30). Deletion of the V3 loop therefore exposes the CD4-binding site to the immune system, allowing the immune system to mount a response against this critical region (18). Another advantage of using the mutant HIV-1 gp120 envelope glycoprotein as an immunogen is that it has significantly reduced affinity for cell surface CD4. An efficient humoral immune response depends on the binding of antigen to B cell surface immunoglobulin. The presence of the high-affinity CD4 receptor on large numbers of cells in the body may significantly diminish the ability of native gp120 to induce an effective humoral immune response. The rationale of mutating gp120 at the CD4 binding site is to redirect the mutant HIV-1 gp120 envelope glycoprotein away from cell surface CD4 toward immunoglobulin-bearing B cells, thereby allowing the immune system to mount a response against, inter alia, the CD4-binding site.

References

1. Klatzmann, D.R., et. al. (1990) Immunodeficiency Reviews 2, 43-66.
- 5 2. Lasky, L.A., et. al. (1987) Cell 50, 975-985.
3. Maddon, P.J., et. al. (1986) Cell 47, 333-348.
- 10 4. Maddon, P.J., et. al. (1988) Cell 54, 865-874.
5. Maddon, P.J., et. al. (1985) Cell 42, 93-104.
- 15 6. Maddon, P.J., et. al. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 9155-9159.
7. Richardson, N.E., et. al. (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 6102-6106.
- 20 8. Chao, B.H., et. al. (1989) J. Biol. Chem. 264, 5812-5817.
9. Arthos, J., et. al. (1989) Cell 57, 469-481.
- 25 10. Wang, J., et. al. (1990) Nature 348, 411-418.
11. Ryu, S.-E., et. al. (1990) Nature 348, 419-426.
12. Leonard, C.K., et. al. (1990) J. Biol. Chem. 265, 10373-10382.
- 30 13. Earl, P.L., et. al. (1990) Proc. Natl. Acad. Sci. U.S.A. 87, 648-652.
- 35 14. Helseth, E., et. al. (1991) J. Virol. 65, 2119-2123.

15. Bolognesi, D.P. (1990) TIBTech 8, 40-45.
16. Olshevsky, U., et. al. (1990) J. Virol. 64, 5701-5707.
- 5 17. Steimer, K.S., et. al. (1991) AIDS 5, S135-143.
18. Wyatt, R., et. al. (1992) J. Virol. 66, 6997-7004.
19. Zolla-Pazner, S., et. al. (1992) Sem. in Virology 3,  
10 203-211.
20. Steimer, K.S., et. al. (1991) Science 254, 105-108.
21. Pollard, S.R., et. al. (1992) EMBO J. 11, 585-591.
- 15 22. Okayama, H. (1983) Mol. Cell. Biol. 3, 280-289.
23. Maniatis, T., et. al. (1990) Molecular Cloning, Vol. 1-  
3.
- 20 24. Thomsen, D.R., et. al. (1984) Proc. Natl. Acad. Sci.  
U.S.A. 81, 659-663.
- 25 25. Pennica, D., et. al. (1983) Nature 301, 214-221.
26. Wain-Hobson, S., et. al. (1985) Cell 40, 9-17.
27. Koyanagi, Y. (1987) Science 236, 819-822.
- 30 28. Reed, L.J. (1938) Am. J. Hyg., 27, 493-497.
29. Cohn, E.J. et al., (1944) J. Clin. Invest. 23, 417-432.
30. Shiow-Her, C., et al. (1992) J. of Cellular Biochem.,  
35 Supplement 16E, Abstrtact Q105.

(j) **APPLICANT:** Progenics Pharmaceuticals, Inc.

(ii) TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES THEREOF

(iii) NUMBER OF SEQUENCES: 29

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Cooper & Dunham  
(B) STREET: 30 Rockefeller Plaza  
(C) CITY: New York  
(D) STATE: New York  
(E) COUNTRY: USA  
(F) ZIP: 10112

(v) **COMPUTER READABLE FORM:**

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.24

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:  
(B) FILING DATE:  
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/037,816  
(B) FILING DATE: 26-MAR-1993

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: White, John P.  
(B) REGISTRATION NUMBER: 28,678  
(C) REFERENCE/DOCKET NUMBER: 41190-A-PCT/JPW/AJM

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (212) 977-9550  
(B) TELEFAX: (212) 664-0525  
(C) TELEX: 422523 COOPU

(2) INFORMATION FOR SEQ ID NO:1:

**(i) SEQUENCE CHARACTERISTICS:**

(A) LENGTH: 45 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Xaa Xaa Xaa Cys Xaa Ile Xaa Xaa Xaa Xaa Xaa Xaa Trp Xaa Xaa Xaa  
1 5 10 15

Xaa Xaa Ala Xaa Tyr Xaa Xaa Pro Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa  
20 25 30

Ser Xaa Xaa Thr Gly Xaa Xaa Xaa Xaa Arg Xaa Gly Xaa

54

35

40

45

## (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 45 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

Thr Leu Pro Cys Arg Ile Lys Gln Phe Ile Asn Met Trp Gln Glu Val
 1             5             10             15
Gly Lys Ala Met Tyr Ala Pro Pro Ile Ser Gly Gln Ile Arg Cys Ser
      20             25             30
Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly
      35             40             45

```

## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 45 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

Thr Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val
 1             5             10             15
Gly Lys Ala Met Tyr Ala Pro Pro Ile Arg Gly Gln Ile Arg Cys Ser
      20             25             30
Ser Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly
      35             40             45

```

## (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 45 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GATCCTGCAG TCACCGTCCT TGACACGATG GATGCAATGA AGAGA

45

## (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 36 base pairs  
 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AAGTCTTCTC CTCGGTCTTG TCTTTTAAAC ACCCAG

36

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 36 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TTCAGAAGAG GAGCCAGAAC AGAAAAATTG TGGGTC

36

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GGAAAAAAGC GGCCGTCAT TTTCTCTCT GCACCACTC

39

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GATCGGCGCC AGAGTAGAAA AGTTGTGGGT CAC

33

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 49 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

SUBSTITUTE SHEET (RULE 26)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CTGTAGAAAT TAATTGTACA GGTGCTGGAC ATTGTAACAT TAGTAGAGC

49

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTCGAGCATG CATTGAAGC TCGCTGATC

29

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CAATTATATA ACATGGTGCA GGAAGTAGG

29

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CAAATTATATA ACATGGTGCA GGAAGTAGG

29

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3125 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1555..3115
- (D) OTHER INFORMATION:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

|   |      |
|---|------|
| TTGACATTGA TTATTGACTA GTTATTAATA GTAATCAATT ACGGGGTCAT TAGTTCATAG | 60   |
| CCCATATATG GAGTTCGCG TTACATAACT TACGGTAAAT GGCCCGCCTG GCTGACCGCC  | 120  |
| CAACGACCCC CGCCCATGGA CGTCAATAAT GACGTATGTT CCCATAGTAA CGCCAATAGG | 180  |
| GACTTTCCAT TGACGTCAT GGGTGGACTA TTTACGGTAA ACTGCCCACT TGGCAGTACA  | 240  |
| TCAAGTGTAT CATATGCCAA GTACGCCCC TATTGACGTC AATGACGGTA AATGGCCCGC  | 300  |
| CTGGCATTAT GCCCAGTACA TGACCTTATG GGACTTTCCT ACTTGGCAGT ACATCTACGT | 360  |
| ATTAGTCATC GCTATTACCA TGGTGATGCG GTTTTGGCAG TACATCAATG GGCGTGGATA | 420  |
| GCGGTTTGAC TCACGGGGAT TTCCAAGTCT CCACCCCATT GACGTCAATG GGAGTTTGTT | 480  |
| TTGGCACCAA AATCAACGGG ACTTTCCAAA ATGTCGTAAC AACTCCGCCC CATTGACGCA | 540  |
| AATGGGCGGT AGGCGGTGAC GGTGGGAGGT CTATATAAGC AGAGCTCGTT TAGTGAACCG | 600  |
| TCAGATCGCC TGGAGACGCC ATCCACGCTG TTTTGACCTC CATAGAAGAC ACCGGGACCG | 660  |
| ATCCAGCCTC CGCGGCGGGG AACGGTGCAT TGGAACGCGG ATTCCCCGTG CCAAGAGTGA | 720  |
| CGTAAGTACC GCCTATAGAC TCTATAGGCA CACCCCTTTG GCTCTTATGC ATGCTATACT | 780  |
| GTTTTTGCT TGGGCCAACA CCCCCTCCTA GATAGGTGAT GGTATAGCTT AGCCTATAGG  | 840  |
| TGTGGGTTAT TGACCATTAT TGACCACTCC CCTATTGGTG ACGATACTTT CCATTACTAA | 900  |
| TCCATAACAT GGCCGCTCTT TGCCACAAC ATCTCTATTG GCTATATGCC AATACTCTGT  | 960  |
| CCTTCAGAGA CTGACACGGA CTCTGTATTT TTACAGGATG GGGTCCCAT TATTATTTAC  | 1020 |
| AAATTCACAT ATACAACAAC GCCGTCCCC GTGCCCGCAG TTTTATTAA CATGCGGGAT   | 1080 |
| CTCCACGCGA ATCTCGGGTA CGTGTTCGG ACATGGGCTC TTCTCCGGTA GCGGCGGAGC  | 1140 |
| TCCACATCCG AGCCTGTCCC ATGCCATGC CTCCAGCGGC TCATGGTCGC TCGGCAGCTC  | 1200 |
| CTTGCTCCTA ACAGTGGAGG CCAGACTTAG GCACAGGACA ATGCCACCA CCACCACTGT  | 1260 |
| GCCGCACAAG GCCGTGGCG TAGGGTATGT GTCTGAAAT GAGCTCGGAG ATTGGGCTCG   | 1320 |
| CACCGCTGAC GCAGATGGAA GACTTAAGGC AGCGGCAGAA GAAGATGCAG GCAGCTGAGT | 1380 |
| TGTTGTATTC TGTAGAGTTG GAGGTAAC TCCTTGCGGT GCTGTTAACG GTGGAGGGCA   | 1440 |
| GTGTAGTCTG AGCAGTACTC GTTGCTGCCG CGCGCGCCAC CAGACATAAT AGCTGACAGA | 1500 |
| CTAACAGACT GTTCCTTTCC ATGGGTCTTT TCTGCAGTCA CCGTCTTGA CACG ATG    | 1557 |
| Met   |      |
| 1   |      |
| GAT GCA ATG AAG AGA GGG CTC TGC TGT GTG CTG CTG CTG TGT GGA GCA   | 1605 |
| Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly Ala   |      |
| 5 10 15   |      |
| GTC TTC GTT TCG CCC AGC CAG GAA ATC CAT GCC CGA TTC AGA AGA GGC   | 1653 |
| Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg Gly   |      |
| 20 25 30  |      |
| GCC AGA ACA GAA AAA TTG TGG GTC ACA GTC TAT TAT GGG GTA CCT GTG   | 1701 |
| Ala Arg Thr Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val   |      |
| 35 40 45  |      |
| TGG AAG GAA GCA ACC ACC ACT CTA TTT TGT GCA TCA GAT GCT AAA GCA   | 1749 |



|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| Trp | Lys | Glu | Ala | Thr | Thr | Thr | Leu | Phe | Cys | Ala | Ser | Asp | Ala | Lys | Ala |      |
| 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     | 65  |      |
| TAT | GAT | ACA | GAG | GTA | CAT | AAT | GTT | TGG | GCC | ACA | CAT | GCC | TGT | GTA | CCC | 1797 |
| Tyr | Asp | Thr | Glu | Val | His | Asn | Val | Trp | Ala | Thr | His | Ala | Cys | Val | Pro |      |
|     |     |     | 70  |     |     |     |     | 75  |     |     |     |     |     | 80  |     |      |
| ACA | GAC | CCC | AAC | CCA | CAA | GAA | GTA | GTA | TTG | GTA | AAT | GTC | ACA | GAA | AAT | 1845 |
| Thr | Asp | Pro | Asn | Pro | Gln | Glu | Val | Val | Leu | Val | Asn | Val | Thr | Glu | Asn |      |
|     |     |     | 85  |     |     |     | 90  |     |     |     |     |     | 95  |     |     |      |
| TTT | AAC | ATG | TGG | AAA | AAT | GAC | ATG | GTA | GAA | CAG | ATG | CAT | GAG | GAT | ATA | 1893 |
| Phe | Asn | Met | Trp | Lys | Asn | Asp | Met | Val | Glu | Gln | Met | His | Glu | Asp | Ile |      |
|     | 100 |     |     |     |     | 105 |     |     |     |     |     | 110 |     |     |     |      |
| ATC | AGT | TTA | TGG | GAT | CAA | AGC | CTA | AAG | CCA | TGT | GTA | AAA | TTA | ACC | CCA | 1941 |
| Ile | Ser | Leu | Trp | Asp | Gln | Ser | Leu | Lys | Pro | Cys | Val | Lys | Leu | Thr | Pro |      |
|     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |     |      |
| CTC | TGT | GTT | AGT | TTA | AAG | TGC | ACT | GAT | TTG | GGG | AAT | GCT | ACT | AAT | ACC | 1989 |
| Leu | Cys | Val | Ser | Leu | Lys | Cys | Thr | Asp | Leu | Gly | Asn | Ala | Thr | Asn | Thr |      |
| 130 |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     | 145 |     |      |
| AAT | AGT | AGT | AAT | ACC | AAT | AGT | AGT | AGC | GGG | GAA | ATG | ATG | ATG | GAG | AAA | 2037 |
| Asn | Ser | Ser | Asn | Thr | Asn | Ser | Ser | Ser | Gly | Glu | Met | Met | Met | Glu | Lys |      |
|     |     |     | 150 |     |     |     |     |     | 155 |     |     |     |     | 160 |     |      |
| GGA | GAG | ATA | AAA | AAC | TGC | TCT | TTC | AAT | ATC | AGC | ACA | AGC | ATA | AGA | GGT | 2085 |
| Gly | Glu | Ile | Lys | Asn | Cys | Ser | Phe | Asn | Ile | Ser | Thr | Ser | Ile | Arg | Gly |      |
|     |     | 165 |     |     |     |     | 170 |     |     |     |     |     | 175 |     |     |      |
| AAG | GTG | CAG | AAA | GAA | TAT | GCA | TTT | TTT | TAT | AAA | CTT | GAT | ATA | ATA | CCA | 2133 |
| Lys | Val | Gln | Lys | Glu | Tyr | Ala | Phe | Phe | Tyr | Lys | Leu | Asp | Ile | Ile | Pro |      |
|     | 180 |     |     |     |     | 185 |     |     |     |     |     | 190 |     |     |     |      |
| ATA | GAT | AAT | GAT | ACT | ACC | AGC | TAT | ACG | TTG | ACA | AGT | TGT | AAC | ACC | TCA | 2181 |
| Ile | Asp | Asn | Asp | Thr | Thr | Ser | Tyr | Thr | Leu | Thr | Ser | Cys | Asn | Thr | Ser |      |
|     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |     |      |
| GTC | ATT | ACA | CAG | GCC | TGT | CCA | AAG | GTA | TCC | TTT | GAG | CCA | ATT | CCC | ATA | 2229 |
| Val | Ile | Thr | Gln | Ala | Cys | Pro | Lys | Val | Ser | Phe | Glu | Pro | Ile | Pro | Ile |      |
| 210 |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     | 225 |     |      |
| CAT | TAT | TGT | GCC | CCG | GCT | GGT | TTT | GCG | ATT | CTA | AAA | TGT | AAT | AAT | AAG | 2277 |
| His | Tyr | Cys | Ala | Pro | Ala | Gly | Phe | Ala | Ile | Leu | Lys | Cys | Asn | Asn | Lys |      |
|     |     |     | 230 |     |     |     | 235 |     |     |     |     |     |     | 240 |     |      |
| ACG | TTC | AAT | GGA | ACA | GGA | CCA | TGT | ACA | AAT | GTC | AGC | ACA | GTA | CAA | TGT | 2325 |
| Thr | Phe | Asn | Gly | Thr | Gly | Pro | Cys | Thr | Asn | Val | Ser | Thr | Val | Gln | Cys |      |
|     |     | 245 |     |     |     |     | 250 |     |     |     |     |     | 255 |     |     |      |
| ACA | CAT | GGA | ATT | AGG | CCA | GTA | GTA | TCA | ACT | CAA | CTG | CTG | TTG | AAT | GGC | 2373 |
| Thr | His | Gly | Ile | Arg | Pro | Val | Val | Ser | Thr | Gln | Leu | Leu | Leu | Asn | Gly |      |
|     |     | 260 |     |     |     | 265 |     |     |     |     |     |     | 270 |     |     |      |
| AGT | CTA | GCA | GAA | GAA | GAG | GTA | GTA | ATT | AGA | TCT | GCC | AAT | TTC | ACA | GAC | 2421 |
| Ser | Leu | Ala | Glu | Glu | Glu | Val | Val | Ile | Arg | Ser | Ala | Asn | Phe | Thr | Asp |      |
|     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |     |      |
| AAT | GCT | AAA | ACC | ATA | ATA | GTA | CAG | CTG | AAC | CAA | TCT | GTA | GAA | ATT | AAT | 2469 |
| Asn | Ala | Lys | Thr | Ile | Ile | Val | Gln | Leu | Asn | Gln | Ser | Val | Glu | Ile | Asn |      |
| 290 |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     | 305 |     |      |
| TGT | ACA | AGA | CCC | AAC | AAC | AAT | ACA | AGA | AAA | AGT | ATC | CGT | ATC | CAG | AGG | 2517 |
| Cys | Thr | Arg | Pro | Asn | Asn | Asn | Thr | Arg | Lys | Ser | Ile | Arg | Ile | Gln | Arg |      |
|     |     |     | 310 |     |     |     | 315 |     |     |     |     |     | 320 |     |     |      |
| GGA | CCA | GGG | AGA | GCA | TTT | GTT | ACA | ATA | GGA | AAA | ATA | GGA | AAT | ATG | AGA | 2565 |
| Gly | Pro | Gly | Arg | Ala | Phe | Val | Thr | Ile | Gly | Lys | Ile | Gly | Asn | Met | Arg |      |

59

| 325   | 330 | 335 |      |
|---|-----|-----|------|
| CAA GCA CAT TGT AAC ATT AGT AGA GCA AAA TGG AAT GCC ACT TTA AAA<br>Gln Ala His Cys Asn Ile Ser Arg Ala Lys Trp Asn Ala Thr Leu Lys<br>340 345 350     |     |     | 2613 |
| CAG ATA GCT AGC AAA TTA AGA GAA CAA TTT GGA AAT AAT AAA ACA ATA<br>Gln Ile Ala Ser Lys Leu Arg Glu Gln Phe Gly Asn Asn Lys Thr Ile<br>355 360 365     |     |     | 2661 |
| ATC TTT AAG CAA TCC TCA GGA GGG GAC CCA GAA ATT GTA ACG CAC AGT<br>Ile Phe Lys Gln Ser Ser Gly Gly Asp Pro Glu Ile Val Thr His Ser<br>370 375 380 385 |     |     | 2709 |
| TTT AAT TGT GGA GGG GAA TTT TTC TAC TGT AAT TCA ACA CAA CTG TTT<br>Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn Ser Thr Gln Leu Phe<br>390 395 400     |     |     | 2757 |
| AAT AGT ACT TGG TTT AAT AGT ACT TGG AGT ACT GAA GGG TCA AAT AAC<br>Asn Ser Thr Trp Phe Asn Ser Thr Trp Ser Thr Glu Gly Ser Asn Asn<br>405 410 415     |     |     | 2805 |
| ACT GAA GGA AGT GAC ACA ATC ACA CTC CCA TGC AGA ATA AAA CAA TTT<br>Thr Glu Gly Ser Asp Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln Phe<br>420 425 430     |     |     | 2853 |
| ATA AAC ATG TGG CAG GAA GTA GGA AAA GCA ATG TAT GCC CCT CCC ATC<br>Ile Asn Met Trp Gln Glu Val Gly Lys Ala Met Tyr Ala Pro Pro Ile<br>435 440 445     |     |     | 2901 |
| AGC GGA CAA ATT AGA TGT TCA TCA AAT ATT ACA GGG CTG CTA TTA ACA<br>Ser Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu Leu Thr<br>450 455 460 465 |     |     | 2949 |
| AGA GAT GGT GGT AAT AAC AAC AAT GGG TCC GAG ATC TTC AGA CCT GGA<br>Arg Asp Gly Gly Asn Asn Asn Asn Gly Ser Glu Ile Phe Arg Pro Gly<br>470 475 480     |     |     | 2997 |
| GGA GGA GAT ATG AGG GAC AAT TGG AGA AGT GAA TTA TAT AAA TAT AAA<br>Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr Lys<br>485 490 495     |     |     | 3045 |
| GTA GTA AAA ATT GAA CCA TTA GGA GTA GCA CCC ACC AAG GCA AAG AGA<br>Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys Arg<br>500 505 510     |     |     | 3093 |
| AGA GTG GTG CAG AGA GAA AAA T GAGCGGCCGC<br>Arg Val Val Gln Arg Glu Lys<br>515 520  |     |     | 3125 |

## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 520 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly  
1 5 10 15

Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg  
20 25 30

Gly Ala Arg Thr Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro

60

|   |     |     |     |
|---|-----|-----|-----|
| 35  | 40  | 45  |     |
| Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys |     |     |     |
| 50  | 55  | 60  |     |
| Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val |     |     |     |
| 65  | 70  | 75  | 80  |
| Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Val Asn Val Thr Glu |     |     |     |
|   | 85  | 90  | 95  |
| Asn Phe Asn Met Trp Lys Asn Asp Met Val Glu Gln Met His Glu Asp |     |     |     |
|   | 100 | 105 | 110 |
| Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr |     |     |     |
|   | 115 | 120 | 125 |
| Pro Leu Cys Val Ser Leu Lys Cys Thr Asp Leu Gly Asn Ala Thr Asn |     |     |     |
|   | 130 | 135 | 140 |
| Thr Asn Ser Ser Asn Thr Asn Ser Ser Ser Gly Glu Met Met Met Glu |     |     |     |
| 145   | 150 | 155 | 160 |
| Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn Ile Ser Thr Ser Ile Arg |     |     |     |
|   | 165 | 170 | 175 |
| Gly Lys Val Gln Lys Glu Tyr Ala Phe Phe Tyr Lys Leu Asp Ile Ile |     |     |     |
|   | 180 | 185 | 190 |
| Pro Ile Asp Asn Asp Thr Thr Ser Tyr Thr Leu Thr Ser Cys Asn Thr |     |     |     |
|   | 195 | 200 | 205 |
| Ser Val Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro |     |     |     |
|   | 210 | 215 | 220 |
| Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asn |     |     |     |
| 225   | 230 | 235 | 240 |
| Lys Thr Phe Asn Gly Thr Gly Pro Cys Thr Asn Val Ser Thr Val Gln |     |     |     |
|   | 245 | 250 | 255 |
| Cys Thr His Gly Ile Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn |     |     |     |
|   | 260 | 265 | 270 |
| Gly Ser Leu Ala Glu Glu Glu Val Val Ile Arg Ser Ala Asn Phe Thr |     |     |     |
|   | 275 | 280 | 285 |
| Asp Asn Ala Lys Thr Ile Ile Val Gln Leu Asn Gln Ser Val Glu Ile |     |     |     |
|   | 290 | 295 | 300 |
| Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gln |     |     |     |
| 305   | 310 | 315 | 320 |
| Arg Gly Pro Gly Arg Ala Phe Val Thr Ile Gly Lys Ile Gly Asn Met |     |     |     |
|   | 325 | 330 | 335 |
| Arg Gln Ala His Cys Asn Ile Ser Arg Ala Lys Trp Asn Ala Thr Leu |     |     |     |
|   | 340 | 345 | 350 |
| Lys Gln Ile Ala Ser Lys Leu Arg Glu Gln Phe Gly Asn Asn Lys Thr |     |     |     |
|   | 355 | 360 | 365 |
| Ile Ile Phe Lys Gln Ser Ser Gly Gly Asp Pro Glu Ile Val Thr His |     |     |     |
|   | 370 | 375 | 380 |
| Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn Ser Thr Gln Leu |     |     |     |
| 385   | 390 | 395 | 400 |
| Phe Asn Ser Thr Trp Phe Asn Ser Thr Trp Ser Thr Glu Gly Ser Asn |     |     |     |

61

405                      410                      415  
 Asn Thr Glu Gly Ser Asp Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln  
                          420                      425                      430  
 Phe Ile Asn Met Trp Gln Glu Val Gly Lys Ala Met Tyr Ala Pro Pro  
                          435                      440                      445  
 Ile Ser Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu Leu  
                          450                      455                      460  
 Thr Arg Asp Gly Gly Asn Asn Asn Asn Gly Ser Glu Ile Phe Arg Pro  
                          465                      470                      475                      480  
 Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr  
                          485                      490                      495  
 Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys  
                          500                      505                      510  
 Arg Arg Val Val Gln Arg Glu Lys  
                          515                      520

## (2) INFORMATION FOR SEQ ID NO:15:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1532 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 1..1522  
 (D) OTHER INFORMATION:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

|  |     |
|--|-----|
| ATG GAT GCA ATG AAG AGA GGG CTC TGC TGT GTG CTG CTG CTG TGT GGA            | 48  |
| Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly            |     |
| 1                      5                      10                      15   |     |
| GCA GTC TTC GTT TCG CCC AGC CAG GAA ATC CAT GCC CGA TTC AGA AGA            | 96  |
| Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg            |     |
| 20                      25                      30                         |     |
| GGC GGC AGA GTA GAA AAG TTG TGG GTC ACA GTC TAT TAT GGG GTA CCT            | 144 |
| Gly Gly Arg Val Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro            |     |
| 35                      40                      45                         |     |
| GTG TGG AAA GAA GCA ACC ACC ACT CTA TTT TGT GCA TCA GAT GCT AAA            | 192 |
| Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys            |     |
| 50                      55                      60                         |     |
| GCA TAT GAT ACA GAG GTA CAT AAT GTT TGG GCC ACA CAT GCC TGT GTA            | 240 |
| Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val            |     |
| 65                      70                      75                      80 |     |
| CCC ACA GAC CCC AAC CCA CAA GAA GTA GTA TTG GAA AAT GTA ACA GAA            | 288 |
| Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Glu Asn Val Thr Glu            |     |
| 85                      90                      95                         |     |
| CAT TTT AAC ATG TGG AAA AAT AAC ATG GTA GAA CAG ATG CAG GAG GAT            | 336 |
| His Phe Asn Met Trp Lys Asn Asn Met Val Glu Gln Met Gln Glu Asp            |     |
| 100                      105                      110                      |     |

62

|   |      |
|---|------|
| ATA ATC AGT TTA TGG GAT CAA AGC CTA AAG CCA TGT GTA AAA TTA ACC<br>Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr<br>115 120 125     | 384  |
| CCA CTC TGT GTT ACT TTA AAT TGC AAG GAT GTG AAT GCT ACT AAT ACC<br>Pro Leu Cys Val Thr Leu Asn Cys Lys Asp Val Asn Ala Thr Asn Thr<br>130 135 140     | 432  |
| ACT AAT GAT AGC GAG GGA ACG ATG GAG AGA GGA GAA ATA AAA AAC TGC<br>Thr Asn Asp Ser Glu Gly Thr Met Glu Arg Gly Glu Ile Lys Asn Cys<br>145 150 155 160 | 480  |
| TCT TTC AAT ATC ACC ACA AGC ATA AGA GAT GAG GTG CAG AAA GAA TAT<br>Ser Phe Asn Ile Thr Thr Ser Ile Arg Asp Glu Val Gln Lys Glu Tyr<br>165 170 175     | 528  |
| GCT CTT TTT TAT AAA CTT GAT GTA GTA CCA ATA GAT AAT AAT AAT ACC<br>Ala Leu Phe Tyr Lys Leu Asp Val Val Pro Ile Asp Asn Asn Asn Thr<br>180 185 190     | 576  |
| AGC TAT AGG TTG ATA AGT TGT GAC ACC TCA GTC ATT ACA CAG GCC TGT<br>Ser Tyr Arg Leu Ile Ser Cys Asp Thr Ser Val Ile Thr Gln Ala Cys<br>195 200 205     | 624  |
| CCA AAG ATA TCC TTT GAG CCA ATT CCC ATA CAT TAT TGT GCC CCG GCT<br>Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala<br>210 215 220     | 672  |
| GGT TTT GCG ATT CTA AAG TGT AAT GAT AAG ACG TTC AAT GGA AAA GGA<br>Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Thr Phe Asn Gly Lys Gly<br>225 230 235 240 | 720  |
| CCA TGT AAA AAT GTC AGC ACA GTA CAA TGT ACA CAT GGA ATT AGG CCA<br>Pro Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro<br>245 250 255     | 768  |
| GTA GTA TCA ACT CAA CTG CTG CTA AAT GGC AGT CTA GCA GAA GAA GAG<br>Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu<br>260 265 270     | 816  |
| GTA GTA ATT AGA TCT GAC AAT TTC ACG AAC AAT GCT AAA ACC ATA ATA<br>Val Val Ile Arg Ser Asp Asn Phe Thr Asn Asn Ala Lys Thr Ile Ile<br>275 280 285     | 864  |
| GTA CAG CTG AAA GAA TCT GTA GAA ATT AAT TGT ACA AGA CCC AAC AAC<br>Val Gln Leu Lys Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn<br>290 295 300     | 912  |
| AAT ACA AGA AAA AGT ATA CAT ATA GGA CCA GGG AGA GCA TTT TAT ACT<br>Asn Thr Arg Lys Ser Ile His Ile Gly Pro Gly Arg Ala Phe Tyr Thr<br>305 310 315 320 | 960  |
| ACA GGA GAA ATA ATA GGA GAT ATA AGA CAA GCA CAT TGT AAC ATT AGT<br>Thr Gly Glu Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Ile Ser<br>325 330 335     | 1008 |
| AGA GCA AAA TGG AAT GAC ACT TTA AAA CAG ATA GTT ATA AAA TTA AGA<br>Arg Ala Lys Trp Asn Asp Thr Leu Lys Gln Ile Val Ile Lys Leu Arg<br>340 345 350     | 1056 |
| GAA CAA TTT GAG AAT AAA ACA ATA GTC TTT AAT CAC TCC TCA GGA GGG<br>Glu Gln Phe Glu Asn Lys Thr Ile Val Phe Asn His Ser Ser Gly Gly<br>355 360 365     | 1104 |
| GAC CCA GAA ATT GTA ATG CAC AGT TTT AAT TGT GGA GGA GAA TTT TTC<br>Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu Phe Phe<br>370 375 380     | 1152 |
| TAC TGT AAT TCA ACA CAA CTG TTT AAT AGT ACT TGG AAT AAT AAT ACT   | 1200 |

63

|   |      |
|---|------|
| Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp Asn Asn Asn Thr |      |
| 385 390 395 400   |      |
| GAA GGG TCA AAT AAC ACT GAA GGA AAT ACT ATC ACA CTC CCA TGC AGA | 1248 |
| Glu Gly Ser Asn Asn Thr Glu Gly Asn Thr Ile Thr Leu Pro Cys Arg |      |
| 405 410 415   |      |
| ATA AAA CAA ATT ATA AAC ATG TGG CAG GAA GTA GGA AAA GCA ATG TAT | 1296 |
| Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Lys Ala Met Tyr |      |
| 420 425 430   |      |
| GCC CCT CCC ATC AGA GGA CAA ATT AGA TGT TCA TCA AAT ATT ACA GGG | 1344 |
| Ala Pro Pro Ile Arg Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly |      |
| 435 440 445   |      |
| CTG CTA TTA ACA AGA GAT GGT GGT ATT AAT GAG AAT GGG ACC GAG ATC | 1392 |
| Leu Leu Leu Thr Arg Asp Gly Ile Asn Glu Asn Gly Thr Glu Ile     |      |
| 450 455 460   |      |
| TTC AGA CCT GGA GGA GGA GAT ATG AGG GAC AAT TGG AGA AGT GAA TTA | 1440 |
| Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu |      |
| 465 470 475 480   |      |
| TAT AAA TAT AAA GTA GTA AAA ATT GAA CCA TTA GGA GTA GCA CCC ACC | 1488 |
| Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr |      |
| 485 490 495   |      |
| AAG GCA AAG AGA AGA GTG GTG CAA AGA GAA AAA T GAGCGGCCGC        | 1532 |
| Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys                     |      |
| 500 505   |      |

## (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 507 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

|   |  |
|---|--|
| Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly |  |
| 1 5 10 15   |  |
| Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg |  |
| 20 25 30  |  |
| Gly Gly Arg Val Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro |  |
| 35 40 45  |  |
| Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys |  |
| 50 55 60  |  |
| Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val |  |
| 65 70 75 80   |  |
| Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Glu Asn Val Thr Glu |  |
| 85 90 95  |  |
| His Phe Asn Met Trp Lys Asn Asn Met Val Glu Gln Met Gln Glu Asp |  |
| 100 105 110   |  |
| Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr |  |
| 115 120 125   |  |
| Pro Leu Cys Val Thr Leu Asn Cys Lys Asp Val Asn Ala Thr Asn Thr |  |
| 130 135 140   |  |

SUBSTITUTE SHEET (RULE 26)

64

Thr Asn Asp Ser Glu Gly Thr Met Glu Arg Gly Glu Ile Lys Asn Cys  
 145 150 155 160  
 Ser Phe Asn Ile Thr Thr Ser Ile Arg Asp Glu Val Gln Lys Glu Tyr  
 165 170 175  
 Ala Leu Phe Tyr Lys Leu Asp Val Val Pro Ile Asp Asn Asn Asn Thr  
 180 185 190  
 Ser Tyr Arg Leu Ile Ser Cys Asp Thr Ser Val Ile Thr Gln Ala Cys  
 195 200 205  
 Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala  
 210 215 220  
 Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Thr Phe Asn Gly Lys Gly  
 225 230 235 240  
 Pro Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro  
 245 250 255  
 Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu  
 260 265 270  
 Val Val Ile Arg Ser Asp Asn Phe Thr Asn Asn Ala Lys Thr Ile Ile  
 275 280 285  
 Val Gln Leu Lys Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn  
 290 295 300  
 Asn Thr Arg Lys Ser Ile His Ile Gly Pro Gly Arg Ala Phe Tyr Thr  
 305 310 315 320  
 Thr Gly Glu Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Ile Ser  
 325 330 335  
 Arg Ala Lys Trp Asn Asp Thr Leu Lys Gln Ile Val Ile Lys Leu Arg  
 340 345 350  
 Glu Gln Phe Glu Asn Lys Thr Ile Val Phe Asn His Ser Ser Gly Gly  
 355 360 365  
 Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu Phe Phe  
 370 375 380  
 Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp Asn Asn Asn Thr  
 385 390 395 400  
 Glu Gly Ser Asn Asn Thr Glu Gly Asn Thr Ile Thr Leu Pro Cys Arg  
 405 410 415  
 Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly Lys Ala Met Tyr  
 420 425 430  
 Ala Pro Pro Ile Arg Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly  
 435 440 445  
 Leu Leu Leu Thr Arg Asp Gly Gly Ile Asn Glu Asn Gly Thr Glu Ile  
 450 455 460  
 Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu  
 465 470 475 480  
 Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr  
 485 490 495  
 Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys  
 500 505

## (2) INFORMATION FOR SEQ ID NO:17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1484 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 1..1474  
 (D) OTHER INFORMATION:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

|   |     |
|---|-----|
| ATG GAT GCA ATG AAG AGA GGG CTC TGC TGT GTG CTG CTG CTG TGT GGA | 48  |
| Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Cys Gly     |     |
| 1 5 10 15   |     |
| GCA GTC TTC GTT TCG CCC AGC CAG GAA ATC CAT GCC CGA TTC AGA AGA | 96  |
| Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg |     |
| 20 25 30  |     |
| GGC GCC AGA ACA GAA AAA TTG TGG GTC ACA GTC TAT TAT GGG GTA CCT | 144 |
| Gly Ala Arg Thr Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro |     |
| 35 40 45  |     |
| GTG TGG AAG GAA GCA ACC ACC ACT CTA TTT TGT GCA TCA GAT GCT AAA | 192 |
| Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys |     |
| 50 55 60  |     |
| GCA TAT GAT ACA GAG GTA CAT AAT GTT TGG GCC ACA CAT GCC TGT GTA | 240 |
| Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val |     |
| 65 70 75 80   |     |
| CCC ACA GAC CCC AAC CCA CAA GAA GTA GTA TTG GTA AAT GTG ACA GAA | 288 |
| Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Val Asn Val Thr Glu |     |
| 85 90 95  |     |
| AAT TTT AAC ATG TGG AAA AAT GAC ATG GTA GAA CAG ATG CAT GAG GAT | 336 |
| Asn Phe Asn Met Trp Lys Asn Asp Met Val Glu Gln Met His Glu Asp |     |
| 100 105 110   |     |
| ATA ATC AGT TTA TGG GAT CAA AGC CTA AAG CCA TGT GTA AAA TTA ACC | 384 |
| Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr |     |
| 115 120 125   |     |
| CCA CTC TGT GTT AGT TTA AAG TGC ACT GAT TTG GGG AAT GCT ACT AAT | 432 |
| Pro Leu Cys Val Ser Leu Lys Cys Thr Asp Leu Gly Asn Ala Thr Asn |     |
| 130 135 140   |     |
| ACC AAT AGT AGT AAT ACC AAT AGT AGT AGC GGG GAA ATG ATG ATG GAG | 480 |
| Thr Asn Ser Ser Asn Thr Asn Ser Ser Ser Gly Glu Met Met Met Glu |     |
| 145 150 155 160   |     |
| AAA GGA GAG ATA AAA AAC TGC TCT TTC AAT ATC AGC ACA AGC ATA AGA | 528 |
| Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn Ile Ser Thr Ser Ile Arg |     |
| 165 170 175   |     |
| GGT AAG GTG CAG AAA GAA TAT GCA TTT TTT TAT AAA CTT GAT ATA ATA | 576 |
| Gly Lys Val Gln Lys Glu Tyr Ala Phe Phe Tyr Lys Leu Asp Ile Ile |     |
| 180 185 190   |     |
| CCA ATA GAT AAT GAT ACT ACC AGC TAT ACG TTG ACA AGT TGT AAC ACC | 624 |
| Pro Ile Asp Asn Asp Thr Thr Ser Tyr Thr Leu Thr Ser Cys Asn Thr |     |
| 195 200 205   |     |



66

|   |      |
|---|------|
| TCA GTC ATT ACA CAG GCC TGT CCA AAG GTA TCC TTT GAG CCA ATT CCC<br>Ser Val Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro<br>210 215 220     | 672  |
| ATA CAT TAT TGT GCC CCG GCT GGT TTT GCG ATT CTA AAA TGT AAT AAT<br>Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asn<br>225 230 235 240 | 720  |
| AAG ACG TTC AAT GGA ACA GGA CCA TGT ACA AAT GTC AGC ACA GTA CAA<br>Lys Thr Phe Asn Gly Thr Gly Pro Cys Thr Asn Val Ser Thr Val Gln<br>245 250 255     | 768  |
| TGT ACA CAT GGA ATT AGG CCA GTA GTA TCA ACT CAA CTG CTG TTG AAT<br>Cys Thr His Gly Ile Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn<br>260 265 270     | 816  |
| GGC AGT CTA GCA GAA GAA GAG GTA GTA ATT AGA TCT GCC AAT TTC ACA<br>Gly Ser Leu Ala Glu Glu Glu Val Val Ile Arg Ser Ala Asn Phe Thr<br>275 280 285     | 864  |
| GAC AAT GCT AAA ACC ATA ATA GTA CAG CTG AAC CAA TCT GTA GAA ATT<br>Asp Asn Ala Lys Thr Ile Ile Val Gln Leu Asn Gln Ser Val Glu Ile<br>290 295 300     | 912  |
| AAT TGT ACA GGT GCT GGA CAT TGT AAC ATT AGT AGA GCA AAA TGG AAT<br>Asn Cys Thr Gly Ala Gly His Cys Asn Ile Ser Arg Ala Lys Trp Asn<br>305 310 315 320 | 960  |
| GCC ACT TTA AAA CAG ATA GCT AGC AAA TTA AGA GAA CAA TTT GGA AAT<br>Ala Thr Leu Lys Gln Ile Ala Ser Lys Leu Arg Glu Gln Phe Gly Asn<br>325 330 335     | 1008 |
| AAT AAA ACA ATA ATC TTT AAG CAA TCC TCA GGA GGG GAC CCA GAA ATT<br>Asn Lys Thr Ile Ile Phe Lys Gln Ser Ser Gly Gly Asp Pro Glu Ile<br>340 345 350     | 1056 |
| GTA ACG CAC AGT TTT AAT TGT GGA GGG GAA TTT TTC TAC TGT AAT TCA<br>Val Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn Ser<br>355 360 365     | 1104 |
| ACA CAA CTG TTT AAT AGT ACT TGG TTT AAT AGT ACT TGG AGT ACT GAA<br>Thr Gln Leu Phe Asn Ser Thr Trp Phe Asn Ser Thr Trp Ser Thr Glu<br>370 375 380     | 1152 |
| GGG TCA AAT AAC ACT GAA GGA AGT GAC ACA ATC ACA CTC CCA TGC AGA<br>Gly Ser Asn Asn Thr Glu Gly Ser Asp Thr Ile Thr Leu Pro Cys Arg<br>385 390 395 400 | 1200 |
| ATA AAA CAA TTT ATA AAC ATG TGG CAG GAA GTA GGA AAA GCA ATG TAT<br>Ile Lys Gln Phe Ile Asn Met Trp Gln Glu Val Gly Lys Ala Met Tyr<br>405 410 415     | 1248 |
| GCC CCT CCC ATC AGC GGA CAA ATT AGA TGT TCA TCA AAT ATT ACA GGG<br>Ala Pro Pro Ile Ser Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly<br>420 425 430     | 1296 |
| CTG CTA TTA ACA AGA GAT GGT GGT AAT AAC AAC AAT GGG TCC GAG ATC<br>Leu Leu Leu Thr Arg Asp Gly Gly Asn Asn Asn Asn Gly Ser Glu Ile<br>435 440 445     | 1344 |
| TTC AGA CCT GGA GGA GGA GAT ATG AGG GAC AAT TGG AGA AGT GAA TTA<br>Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu<br>450 455 460     | 1392 |
| TAT AAA TAT AAA GTA GTA AAA ATT GAA CCA TTA GGA GTA GCA CCC ACC<br>Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr<br>465 470 475 480 | 1440 |
| AAG GCA AAG AGA AGA GTG GTG CAG AGA GAA AAA T GAGCGGCCGC  | 1484 |

Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys  
485 490

## (2) INFORMATION FOR SEQ ID NO:18:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 491 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly  
1 5 10 15  
Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg  
20 25 30  
Gly Ala Arg Thr Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro  
35 40 45  
Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys  
50 55 60  
Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val  
65 70 75 80  
Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Val Asn Val Thr Glu  
85 90 95  
Asn Phe Asn Met Trp Lys Asn Asp Met Val Glu Gln Met His Glu Asp  
100 105 110  
Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr  
115 120 125  
Pro Leu Cys Val Ser Leu Lys Cys Thr Asp Leu Gly Asn Ala Thr Asn  
130 135 140  
Thr Asn Ser Ser Asn Thr Asn Ser Ser Ser Gly Glu Met Met Met Glu  
145 150 155 160  
Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn Ile Ser Thr Ser Ile Arg  
165 170 175  
Gly Lys Val Gln Lys Glu Tyr Ala Phe Phe Tyr Lys Leu Asp Ile Ile  
180 185 190  
Pro Ile Asp Asn Asp Thr Thr Ser Tyr Thr Leu Thr Ser Cys Asn Thr  
195 200 205  
Ser Val Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro  
210 215 220  
Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asn  
225 230 235 240  
Lys Thr Phe Asn Gly Thr Gly Pro Cys Thr Asn Val Ser Thr Val Gln  
245 250 255  
Cys Thr His Gly Ile Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn  
260 265 270  
Gly Ser Leu Ala Glu Glu Glu Val Val Ile Arg Ser Ala Asn Phe Thr  
275 280 285

68

Asp Asn Ala Lys Thr Ile Ile Val Gln Leu Asn Gln Ser Val Glu Ile  
 290 295 300  
 Asn Cys Thr Gly Ala Gly His Cys Asn Ile Ser Arg Ala Lys Trp Asn  
 305 310 315 320  
 Ala Thr Leu Lys Gln Ile Ala Ser Lys Leu Arg Glu Gln Phe Gly Asn  
 325 330 335  
 Asn Lys Thr Ile Ile Phe Lys Gln Ser Ser Gly Gly Asp Pro Glu Ile  
 340 345 350  
 Val Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn Ser  
 355 360 365  
 Thr Gln Leu Phe Asn Ser Thr Trp Phe Asn Ser Thr Trp Ser Thr Glu  
 370 375 380  
 Gly Ser Asn Asn Thr Glu Gly Ser Asp Thr Ile Thr Leu Pro Cys Arg  
 385 390 395 400  
 Ile Lys Gln Phe Ile Asn Met Trp Gln Glu Val Gly Lys Ala Met Tyr  
 405 410 415  
 Ala Pro Pro Ile Ser Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly  
 420 425 430  
 Leu Leu Leu Thr Arg Asp Gly Gly Asn Asn Asn Asn Gly Ser Glu Ile  
 435 440 445  
 Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu  
 450 455 460  
 Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr  
 465 470 475 480  
 Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys  
 485 490

## (2) INFORMATION FOR SEQ ID NO:19:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1448 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1439
- (D) OTHER INFORMATION:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

|   |     |
|---|-----|
| ATG GAT GCA ATG AAG AGA GGG CTC TGC TGT STG CTG CTG CTG TGT GGA | 48  |
| Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly |     |
| 1 5 10 15   |     |
| GCA GTC TTC GTT TCG CCC AGC CAG GAA ATC CAT GCC CGA TTC AGA AGA | 96  |
| Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg |     |
| 20 25 30  |     |
| GGC GGC AGA GTA GAA AAG TTG TGG GTC ACA GTC TAT TAT GGG GTA CCT | 144 |
| Gly Gly Arg Val Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro |     |
| 35 40 45  |     |

|   |      |
|---|------|
| GTG TGG AAA GAA GCA ACC ACC ACT CTA TTT TGT GCA TCA GAT GCT AAA<br>Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys<br>50 55 60        | 192  |
| GCA TAT GAT ACA GAG GTA CAT AAT GTT TGG GCC ACA CAT GCC TGT GTA<br>Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val<br>65 70 75 80     | 240  |
| CCC ACA GAC CCC AAC CCA CAA GAA GTA GTA TTG GAA AAT GTA ACA GAA<br>Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Glu Asn Val Thr Glu<br>85 90 95        | 288  |
| CAT TTT AAC ATG TGG AAA AAT AAC ATG GTA GAA CAG ATG CAG GAG GAT<br>His Phe Asn Met Trp Lys Asn Asn Met Val Glu Gln Met Gln Glu Asp<br>100 105 110     | 336  |
| ATA ATC AGT TTA TGG GAT CAA AGC CTA AAG CCA TGT GTA AAA TTA ACC<br>Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr<br>115 120 125     | 384  |
| CCA CTC TGT GTT ACT TTA AAT TGC AAG GAT GTG AAT GCT ACT AAT ACC<br>Pro Leu Cys Val Thr Leu Asn Cys Lys Asp Val Asn Ala Thr Asn Thr<br>130 135 140     | 432  |
| ACT AAT GAT AGC GAG GGA ACG ATG GAG AGA GGA GAA ATA AAA AAC TGC<br>Thr Asn Asp Ser Glu Gly Thr Met Glu Arg Gly Glu Ile Lys Asn Cys<br>145 150 155 160 | 480  |
| TCT TTC AAT ATC ACC ACA AGC ATA AGA GAT GAG GTG CAG AAA GAA TAT<br>Ser Phe Asn Ile Thr Thr Ser Ile Arg Asp Glu Val Gln Lys Glu Tyr<br>165 170 175     | 528  |
| GCT CTT TTT TAT AAA CTT GAT GTA GTA CCA ATA GAT AAT AAT AAT ACC<br>Ala Leu Phe Tyr Lys Leu Asp Val Val Pro Ile Asp Asn Asn Asn Thr<br>180 185 190     | 576  |
| AGC TAT AGG TTG ATA AGT TGT GAC ACC TCA GTC ATT ACA CAG GCC TGT<br>Ser Tyr Arg Leu Ile Ser Cys Asp Thr Ser Val Ile Thr Gln Ala Cys<br>195 200 205     | 624  |
| CCA AAG ATA TCC TTT GAG CCA ATT CCC ATA CAT TAT TGT GCC CCG GCT<br>Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala<br>210 215 220     | 672  |
| GGT TTT GCG ATT CTA AAG TGT AAT GAT AAG ACG TTC AAT GGA AAA GGA<br>Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Thr Phe Asn Gly Lys Gly<br>225 230 235 240 | 720  |
| CCA TGT AAA AAT GTC AGC ACA GTA CAA TGT ACA CAT GGA ATT AGG CCA<br>Pro Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro<br>245 250 255     | 768  |
| GTA GTA TCA ACT CAA CTG CTG CTA AAT GGC AGT CTA GCA GAA GAA GAG<br>Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu<br>260 265 270     | 816  |
| GTA GTA ATT AGA TCT GAC AAT TTC ACG AAC AAT GCT AAA ACC ATA ATA<br>Val Val Ile Arg Ser Asp Asn Phe Thr Asn Asn Ala Lys Thr Ile Ile<br>275 280 285     | 864  |
| GTA CAG CTG AAA GAA TCT GTA GAA ATT AAT TGT ACA GGT GCT GGA CAT<br>Val Gln Leu Lys Glu Ser Val Glu Ile Asn Cys Thr Gly Ala Gly His<br>290 295 300     | 912  |
| TGT AAC ATT AGT AGA GCA AAA TGG AAT GAC ACT TTA AAA CAG ATA GTT<br>Cys Asn Ile Ser Arg Ala Lys Trp Asn Asp Thr Leu Lys Gln Ile Val<br>305 310 315 320 | 960  |
| ATA AAA TTA AGA GAA CAA TTT GAG AAT AAA ACA ATA GTC TTT AAT CAC   | 1008 |

70

|   |      |
|---|------|
| Ile Lys Leu Arg Glu Gln Phe Glu Asn Lys Thr Ile Val Phe Asn His |      |
| 325 330 335   |      |
| TCC TCA GGA GGG GAC CCA GAA ATT GTA ATG CAC AGT TTT AAT TGT GGA | 1056 |
| Ser Ser Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly |      |
| 340 345 350   |      |
| GGA GAA TTT TTC TAC TGT AAT TCA ACA CAA CTG TTT AAT AGT ACT TGG | 1104 |
| Gly Glu Phe Phe Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp |      |
| 355 360 365   |      |
| AAT AAT AAT ACT GAA GGG TCA AAT AAC ACT GAA GGA AAT ACT ATC ACA | 1152 |
| Asn Asn Asn Thr Glu Gly Ser Asn Asn Thr Glu Gly Asn Thr Ile Thr |      |
| 370 375 380   |      |
| CTC CCA TGC AGA ATA AAA CAA ATT ATA AAC ATG TGG CAG GAA GTA GGA | 1200 |
| Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Trp Gln Glu Val Gly |      |
| 385 390 395 400   |      |
| AAA GCA ATG TAT GCC CCT CCC ATC AGA GGA CAA ATT AGA TGT TCA TCA | 1248 |
| Lys Ala Met Tyr Ala Pro Pro Ile Arg Gly Gln Ile Arg Cys Ser Ser |      |
| 405 410 415   |      |
| AAT ATT ACA GGG CTG CTA TTA ACA AGA GAT GGT GGT ATT AAT GAG AAT | 1296 |
| Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Ile Asn Glu Asn |      |
| 420 425 430   |      |
| GGG ACC GAG ATC TTC AGA CCT GGA GGA GGA GAT ATG AGG GAC AAT TGG | 1344 |
| Gly Thr Glu Ile Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp |      |
| 435 440 445   |      |
| AGA AGT GAA TTA TAT AAA TAT AAA GTA GTA AAA ATT GAA CCA TTA GGA | 1392 |
| Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly |      |
| 450 455 460   |      |
| GTA GCA CCC ACC AAG GCA AAG AGA AGA GTG GTG CAA AGA GAA AAA TG  | 1439 |
| Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys     |      |
| 465 470 475   |      |
| AGCGGCCGC   | 1448 |

## (2) INFORMATION FOR SEQ ID NO:20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 479 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

|   |  |
|---|--|
| Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly |  |
| 1 5 10 15   |  |
| Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg |  |
| 20 25 30  |  |
| Gly Gly Arg Val Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro |  |
| 35 40 45  |  |
| Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys |  |
| 50 55 60  |  |
| Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val |  |
| 65 70 75 80   |  |
| Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Glu Asn Val Thr Glu |  |

85

90

95

**SUBSTITUTE SHEET (RULE 26)**

450                      455                      460  
 Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys  
 465                      470                      475

## (2) INFORMATION FOR SEQ ID NO:21:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1484 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1454
- (D) OTHER INFORMATION:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

|  |     |
|--|-----|
| ATG GAT GCA ATG AAG AGA GGG CTC TGC TGT GTG CTG CTG CTG TGT GGA                | 48  |
| Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly                |     |
| 1                      5                      10                      15       |     |
| GCA GTC TTC GTT TCG CCC AGC CAG GAA ATC CAT GCC CGA TTC AGA AGA                | 96  |
| Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg                |     |
| 20                      25                      30                             |     |
| GGC GCC AGA ACA GAA AAA TTG TGG GTC ACA GTC TAT TAT GGG GTA CCT                | 144 |
| Gly Ala Arg Thr Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro                |     |
| 35                      40                      45                             |     |
| GTG TGG AAG GAA GCA ACC ACC ACT CTA TTT TGT GCA TCA GAT GCT AAA                | 192 |
| Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys                |     |
| 50                      55                      60                             |     |
| GCA TAT GAT ACA GAG GTA CAT AAT GTT TGG GCC ACA CAT GCC TGT GTA                | 240 |
| Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val                |     |
| 65                      70                      75                      80     |     |
| CCC ACA GAC CCC AAC CCA CAA GAA GTA GTA TTG GTA AAT GTG ACA GAA                | 288 |
| Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Val Asn Val Thr Glu                |     |
| 85                      90                      95                             |     |
| AAT TTT AAC ATG TGG AAA AAT GAC ATG GTA GAA CAG ATG CAT GAG GAT                | 336 |
| Asn Phe Asn Met Trp Lys Asn Asp Met Val Glu Gln Met His Glu Asp                |     |
| 100                      105                      110                          |     |
| ATA ATC AGT TTA TGG GAT CAA AGC CTA AAG CCA TGT GTA AAA TTA ACC                | 384 |
| Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr                |     |
| 115                      120                      125                          |     |
| CCA CTC TGT GTT AGT TTA AAG TGC ACT GAT TTG GGG AAT GCT ACT AAT                | 432 |
| Pro Leu Cys Val Ser Leu Lys Cys Thr Asp Leu Gly Asn Ala Thr Asn                |     |
| 130                      135                      140                          |     |
| ACC AAT AGT AGT AAT ACC AAT AGT AGT AGC GGG GAA ATG ATG ATG GAG                | 480 |
| Thr Asn Ser Ser Asn Thr Asn Ser Ser Ser Gly Glu Met Met Met Glu                |     |
| 145                      150                      155                      160 |     |
| AAA GGA GAG ATA AAA AAC TGC TCT TTC AAT ATC AGC ACA AGC ATA AGA                | 528 |
| Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn Ile Ser Thr Ser Ile Arg                |     |
| 165                      170                      175                          |     |
| GGT AAG GTG CAG AAA GAA TAT GCA TTT TTT TAT AAA CTT GAT ATA ATA                | 576 |
| Gly Lys Val Gln Lys Glu Tyr Ala Phe Phe Tyr Lys Leu Asp Ile Ile                |     |

73

| 180   | 185 | 190 |      |
|---|-----|-----|------|
| CCA ATA GAT AAT GAT ACT ACC AGC TAT ACG TTG ACA AGT TGT AAC ACC<br>Pro Ile Asp Asn Asp Thr Thr Ser Tyr Thr Leu Thr Ser Cys Asn Thr<br>195 200 205     |     |     | 624  |
| TCA GTC ATT ACA CAG GCC TGT CCA AAG GTA TCC TTT GAG CCA ATT CCC<br>Ser Val Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro<br>210 215 220     |     |     | 672  |
| ATA CAT TAT TGT GCC CCG GCT GGT TTT GCG ATT CTA AAA TGT AAT AAT<br>Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asn<br>225 230 235 240 |     |     | 720  |
| AAG ACG TTC AAT GGA ACA GGA CCA TGT ACA AAT GTC AGC ACA GTA CAA<br>Lys Thr Phe Asn Gly Thr Gly Pro Cys Thr Asn Val Ser Thr Val Gln<br>245 250 255     |     |     | 768  |
| TGT ACA CAT GGA ATT AGG CCA GTA GTA TCA ACT CAA CTG CTG TTG AAT<br>Cys Thr His Gly Ile Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn<br>260 265 270     |     |     | 816  |
| GGC-AGT CTA GCA GAA GAA GAG GTA GTA ATT AGA TCT GCC AAT TTC ACA<br>Gly Ser Leu Ala Glu Glu Glu Val Val Ile Arg Ser Ala Asn Phe Thr<br>275 280 285     |     |     | 864  |
| GAC AAT GCT AAA ACC ATA ATA GTA CAG CTG AAC CAA TCT GTA GAA ATT<br>Asn Asn Ala Lys Thr Ile Ile Val Gln Leu Asn Gln Ser Val Glu Ile<br>290 295 300     |     |     | 912  |
| AAT TGT ACA GGT GCT GGA CAT TGT AAC ATT AGT AGA GCA AAA TGG AAT<br>Asn Cys Thr Gly Ala Gly His Cys Asn Ile Ser Arg Ala Lys Trp Asn<br>305 310 315 320 |     |     | 960  |
| GCC ACT TTA AAA CAG ATA GCT AGC AAA TTA AGA GAA CAA TTT GGA AAT<br>Ala Thr Leu Lys Gln Ile Ala Ser Lys Leu Arg Glu Gln Phe Gly Asn<br>325 330 335     |     |     | 1008 |
| AAT AAA ACA ATA ATC TTT AAG CAA TCC TCA GGA GGG GAC CCA GAA ATT<br>Asn Lys Thr Ile Ile Phe Lys Gln Ser Ser Gly Gly Asp Pro Glu Ile<br>340 345 350     |     |     | 1056 |
| GTA ACG CAC AGT TTT AAT TGT GGA GGG GAA TTT TTC TAC TGT AAT TCA<br>Val Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn Ser<br>355 360 365     |     |     | 1104 |
| ACA CAA CTG TTT AAT AGT ACT TGG TTT AAT AGT ACT TGG AGT ACT GAA<br>Thr Gln Leu Phe Asn Ser Thr Trp Phe Asn Ser Thr Trp Ser Thr Glu<br>370 375 380     |     |     | 1152 |
| GGG TCA AAT AAC ACT GAA GGA AGT GAC ACA ATC ACA CTC CCA TGC AGA<br>Gly Ser Asn Asn Thr Glu Gly Ser Asp Thr Ile Thr Leu Pro Cys Arg<br>385 390 395 400 |     |     | 1200 |
| ATA AAA CAA TTT ATA AAC ATG GTG CAG GAA GTA GGA AAA GCA ATG TAT<br>Ile Lys Gln Phe Ile Asn Met Val Gln Glu Val Gly Lys Ala Met Tyr<br>405 410 415     |     |     | 1248 |
| GCC CCT CCC ATC AGC GGA CAA ATT AGA TGT TCA TCA AAT ATT ACA GGG<br>Ala Pro Pro Ile Ser Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly<br>420 425 430     |     |     | 1296 |
| CTG CTA TTA ACA AGA GAT GGT GGT AAT AAC AAC AAT GCG TCC GAG ATC<br>Leu Leu Leu Thr Arg Asp Gly Gly Asn Asn Asn Asn Gly Ser Glu Ile<br>435 440 445     |     |     | 1344 |
| TTC AGA CCT GGA GGA GGA GAT ATG AGG GAC AAT TGG AGA AGT GAA TTA<br>Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu<br>450 455 460     |     |     | 1392 |

SUBSTITUTE SHEET (RULE 26)



74

TAT AAA TAT AAA GTA GTA AAA ATT GAA CCA TTA GGA GTA GCA CCC ACC 1440  
 Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr  
 465 470 475 480

AAG GCA AAG AGA AG AGTGGTGCAG AGAGAAAAAT GAGCGGCCGC 1484  
 Lys Ala Lys Arg

## (2) INFORMATION FOR SEQ ID NO:22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 484 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly  
 1 5 10 15  
 Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg  
 20 25 30  
 Gly Ala Arg Thr Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro  
 35 40 45  
 Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys  
 50 55 60  
 Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val  
 65 70 75 80  
 Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Val Asn Val Thr Glu  
 85 90 95  
 Asn Phe Asn Met Trp Lys Asn Asp Met Val Glu Gln Met His Glu Asp  
 100 105 110  
 Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr  
 115 120 125  
 Pro Leu Cys Val Ser Leu Lys Cys Thr Asp Leu Gly Asn Ala Thr Asn  
 130 135 140  
 Thr Asn Ser Ser Asn Thr Asn Ser Ser Ser Gly Glu Met Met Met Glu  
 145 150 155 160  
 Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn Ile Ser Thr Ser Ile Arg  
 165 170 175  
 Gly Lys Val Gln Lys Glu Tyr Ala Phe Phe Tyr Lys Leu Asp Ile Ile  
 180 185 190  
 Pro Ile Asp Asn Asp Thr Thr Ser Tyr Thr Leu Thr Ser Cys Asn Thr  
 195 200 205  
 Ser Val Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro  
 210 215 220  
 Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asn  
 225 230 235 240  
 Lys Thr Phe Asn Gly Thr Gly Pro Cys Thr Asn Val Ser Thr Val Gln  
 245 250 255  
 Cys Thr His Gly Ile Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn

75

260                      265                      270  
 Gly Ser Leu Ala Glu Glu Glu Val Val Ile Arg Ser Ala Asn Phe Thr  
       275                      280                      285  
 Asp Asn Ala Lys Thr Ile Ile Val Gln Leu Asn Gln Ser Val Glu Ile  
       290                      295                      300  
 Asn Cys Thr Gly Ala Gly His Cys Asn Ile Ser Arg Ala Lys Trp Asn  
       305                      310                      315                      320  
 Ala Thr Leu Lys Gln Ile Ala Ser Lys Leu Arg Glu Gln Phe Gly Asn  
                               325                      330                      335  
 Asn Lys Thr Ile Ile Phe Lys Gln Ser Ser Gly Gly Asp Pro Glu Ile  
                               340                      345                      350  
 Val Thr His Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn Ser  
                               355                      360                      365  
 Thr Gln Leu Phe Asn Ser Thr Trp Phe Asn Ser Thr Trp Ser Thr Glu  
       370                      375                      380  
 Gly Ser Asn Asn Thr Glu Gly Ser Asp Thr Ile Thr Leu Pro Cys Arg  
       385                      390                      395                      400  
 Ile Lys Gln Phe Ile Asn Met Val Gln Glu Val Gly Lys Ala Met Tyr  
                               405                      410                      415  
 Ala Pro Pro Ile Ser Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly  
                               420                      425                      430  
 Leu Leu Leu Thr Arg Asp Gly Gly Asn Asn Asn Asn Gly Ser Glu Ile  
                               435                      440                      445  
 Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu  
                               450                      455                      460  
 Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr  
       465                      470                      475                      480  
 Lys Ala Lys Arg

## (2) INFORMATION FOR SEQ ID NO:23:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1448 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 1..1438  
 (D) OTHER INFORMATION:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

|  |    |
|--|----|
| ATG GAT GCA ATG AAG AGA GGG CTC TGC TGT GTG CTG CTG CTG TGT GGA          | 48 |
| Met Asp Ala Met Lys Arg Gly Leu Cys Val Leu Leu Leu Cys Gly              |    |
| 1                      5                      10                      15 |    |
| GCA GTC TTC GTT TCG CCC AGC CAG GAA ATC CAT GCC CGA TTC AGA AGA          | 96 |
| Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg          |    |
| 20                      25                      30                       |    |

76

|   |     |
|---|-----|
| GGC GGC AGA GTA GAA AAG TTG TGG GTC ACA GTC TAT TAT GGG GTA CCT<br>Gly Gly Arg Val Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro<br>35 40 45        | 144 |
| GTG TGG AAA GAA GCA ACC ACC ACT CTA TTT TGT GCA TCA GAT GCT AAA<br>Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys<br>50 55 60        | 192 |
| GCA TAT GAT ACA GAG GTA CAT AAT GTT TGG GCC ACA CAT GCC TGT GTA<br>Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val<br>65 70 75 80     | 240 |
| CCC ACA GAC CCC AAC CCA CAA GAA GTA GTA TTG GAA AAT GTA ACA GAA<br>Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Glu Asn Val Thr Glu<br>85 90 95        | 288 |
| CAT TTT AAC ATG TGG AAA AAT AAC ATG GTA GAA CAG ATG CAG GAG GAT<br>His Phe Asn Met Trp Lys Asn Asn Met Val Glu Gln Met Gln Glu Asp<br>100 105 110     | 336 |
| ATA ATC AGT TTA TGG GAT CAA AGC CTA AAG CCA TGT GTA AAA TTA ACC<br>Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr<br>115 120 125     | 384 |
| CCA CTC TGT GTT ACT TTA AAT TGC AAG GAT GTG AAT GCT ACT AAT ACC<br>Pro Leu Cys Val Thr Leu Asn Cys Lys Asp Val Asn Ala Thr Asn Thr<br>130 135 140     | 432 |
| ACT AAT GAT AGC GAG GGA ACG ATG GAG AGA GGA GAA ATA AAA AAC TGC<br>Thr Asn Asp Ser Glu Gly Thr Met Glu Arg Gly Glu Ile Lys Asn Cys<br>145 150 155 160 | 480 |
| TCT TTC AAT ATC ACC ACA AGC ATA AGA GAT GAG GTG CAG AAA GAA TAT<br>Ser Phe Asn Ile Thr Thr Ser Ile Arg Asp Glu Val Gln Lys Glu Tyr<br>165 170 175     | 528 |
| GCT CTT TTT TAT AAA CTT GAT GTA GTA CCA ATA GAT AAT AAT AAT ACC<br>Ala Leu Phe Tyr Lys Leu Asp Val Val Pro Ile Asp Asn Asn Asn Thr<br>180 185 190     | 576 |
| AGC TAT AGG TTG ATA AGT TGT GAC ACC TCA GTC ATT ACA CAG GCC TGT<br>Ser Tyr Arg Leu Ile Ser Cys Asp Thr Ser Val Ile Thr Gln Ala Cys<br>195 200 205     | 624 |
| CCA AAG ATA TCC TTT GAG CCA ATT CCC ATA CAT TAT TGT GCC CCG GCT<br>Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala<br>210 215 220     | 672 |
| GGT TTT GCG ATT CTA AAG TGT AAT GAT AAG ACG TTC AAT GGA AAA GGA<br>Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Thr Phe Asn Gly Lys Gly<br>225 230 235 240 | 720 |
| CCA TGT AAA AAT GTC AGC ACA GTA CAA TGT ACA CAT GGA ATT AGG CCA<br>Pro Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro<br>245 250 255     | 768 |
| GTA GTA TCA ACT CAA CTG CTG CTA AAT GGC AGT CTA GCA GAA GAA GAG<br>Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu<br>260 265 270     | 816 |
| GTA GTA ATT AGA TCT GAC AAT TTC ACG AAC AAT GCT AAA ACC ATA ATA<br>Val Val Ile Arg Ser Asp Asn Phe Thr Asn Asn Ala Lys Thr Ile Ile<br>275 280 285     | 864 |
| GTA CAG CTG AAA GAA TCT GTA GAA ATT AAT TGT ACA GGT GCT GGA CAT<br>Val Gln Leu Lys Glu Ser Val Glu Ile Asn Cys Thr Gly Ala Gly His<br>290 295 300     | 912 |
| TGT AAC ATT AGT AGA GCA AAA TGG AAT GAC ACT TTA AAA CAG ATA GTT   | 960 |

SUBSTITUTE SHEET (RULE 26)

77

Cys Asn Ile Ser Arg Ala Lys Trp Asn Asp Thr Leu Lys Gln Ile Val  
 305 310 315 320  
 ATA AAA TTA AGA GAA CAA TTT GAG AAT AAA ACA ATA GTC TTT AAT CAC 1008  
 Ile Lys Leu Arg Glu Gln Phe Glu Asn Lys Thr Ile Val Phe Asn His  
 325 330 335  
 TSC TCA GGA GGG GAC CCA GAA ATT GTA ATG CAC AGT TTT AAT TGT GGA 1056  
 Ser Ser Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly  
 340 345 350  
 GGA GAA TTT TTC TAC TGT AAT TCA ACA CAA CTG TTT AAT AGT ACT TGG 1104  
 Gly Glu Phe Phe Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp  
 355 360 365  
 AAT AAT AAT ACT GAA GGG TCA AAT AAC ACT GAA GGA AAT ACT ATC ACA 1152  
 Asn Asn Asn Thr Glu Gly Ser Asn Asn Thr Glu Gly Asn Thr Ile Thr  
 370 375 380  
 CTC CCA TGC AGA ATA AAA CAA ATT ATA AAC ATG GTG CAG GAA GTA GGA 1200  
 Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Val Gln Glu Val Gly  
 385 390 395 400  
 AAA GCA ATG TAT GCC CCT CCC ATC AGA GGA CAA ATT AGA TGT TCA TCA 1248  
 Lys Ala Met Tyr Ala Pro Pro Ile Arg Gly Gln Ile Arg Cys Ser Ser  
 405 410 415  
 AAT ATT ACA GGG CTG CTA TTA ACA AGA GAT GGT GGT ATT AAT GAG AAT 1296  
 Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Ile Asn Glu Asn  
 420 425 430  
 GGG ACC GAG ATC TTC AGA CCT GGA GGA GGA GAT ATG AGG GAC AAT TGG 1344  
 Gly Thr Glu Ile Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp  
 435 440 445  
 AGA AGT GAA TTA TAT AAA TAT AAA GTA GTA AAA ATT GAA CCA TTA GGA 1392  
 Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly  
 450 455 460  
 GTA GCA CCC ACC AAG GCA AAG AGA AGA GTG GTG CAA AGA GAA AAA T 1438  
 Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys  
 465 470 475  
 GAGCGGCCGC 1448

## (2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 479 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly  
 1 5 10 15  
 Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg  
 20 25 30  
 Gly Gly Arg Val Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro  
 35 40 45  
 Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys  
 50 55 60

78

Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val  
 65 70 75 80  
 Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Glu Asn Val Thr Glu  
 85 90 95  
 His Phe Asn Met Trp Lys Asn Asn Met Val Glu Gln Met Gln Glu Asp  
 100 105 110  
 Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr  
 115 120 125  
 Pro Leu Cys Val Thr Leu Asn Cys Lys Asp Val Asn Ala Thr Asn Thr  
 130 135 140  
 Thr Asn Asp Ser Glu Gly Thr Met Glu Arg Gly Glu Ile Lys Asn Cys  
 145 150 155 160  
 Ser Phe Asn Ile Thr Thr Ser Ile Arg Asp Glu Val Gln Lys Glu Tyr  
 165 170 175  
 Ala Leu Phe Tyr Lys Leu Asp Val Val Pro Ile Asp Asn Asn Asn Thr  
 180 185 190  
 Ser Tyr Arg Leu Ile Ser Cys Asp Thr Ser Val Ile Thr Gln Ala Cys  
 195 200 205  
 Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala  
 210 215 220  
 Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Thr Phe Asn Gly Lys Gly  
 225 230 235 240  
 Pro Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro  
 245 250 255  
 Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu  
 260 265 270  
 Val Val Ile Arg Ser Asp Asn Phe Thr Asn Asn Ala Lys Thr Ile Ile  
 275 280 285  
 Val Gln Leu Lys Glu Ser Val Glu Ile Asn Cys Thr Gly Ala Gly His  
 290 295 300  
 Cys Asn Ile Ser Arg Ala Lys Trp Asn Asp Thr Leu Lys Gln Ile Val  
 305 310 315 320  
 Ile Lys Leu Arg Glu Gln Phe Glu Asn Lys Thr Ile Val Phe Asn His  
 325 330 335  
 Ser Ser Gly Gly Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly  
 340 345 350  
 Gly Glu Phe Phe Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp  
 355 360 365  
 Asn Asn Asn Thr Glu Gly Ser Asn Asn Thr Glu Gly Asn Thr Ile Thr  
 370 375 380  
 Leu Pro Cys Arg Ile Lys Gln Ile Ile Asn Met Val Gln Glu Val Gly  
 385 390 395 400  
 Lys Ala Met Tyr Ala Pro Pro Ile Arg Gly Gln Ile Arg Cys Ser Ser  
 405 410 415  
 Asn Ile Thr Gly Leu Leu Leu Thr Arg Asp Gly Gly Ile Asn Glu Asn  
 420 425 430

Gly Thr Glu Ile Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp  
435 440 445

Arg Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly  
450 455 460

Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys  
465 470 475

## (2) INFORMATION FOR SEQ ID NO:25:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1571 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS  
(B) LOCATION: 1..1567  
(D) OTHER INFORMATION:

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

|   |     |
|---|-----|
| ATG GAT GCA ATG AAG AGA GGG CTC TGC TGT GTG CTG CTG CTG TGT GGA | 48  |
| Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly |     |
| 1 5 10 15   |     |
| GCA GTC TTC GTT TCG CCC AGC CAG GAA ATC CAT GCC CGA TTC AGA AGA | 96  |
| Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg |     |
| 20 25 30  |     |
| GGC GCC AGA ACA GAA AAA TTG TGG GTC ACA GTC TAT TAT GGG GTA CCT | 144 |
| Gly Ala Arg Thr Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro |     |
| 35 40 45  |     |
| GTG TGG AAG GAA GCA ACC ACC ACT CTA TTT TGT GCA TCA GAT GCT AAA | 192 |
| Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys |     |
| 50 55 60  |     |
| GCA TAT GAT ACA GAG GTA CAT AAT GTT TGG GCC ACA CAT GCC TGT GTA | 240 |
| Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val |     |
| 65 70 75 80   |     |
| CCC ACA GAC CCC AAC CCA CAA GAA GTA GTA TTG GTA AAT GTG ACA GAA | 288 |
| Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Val Asn Val Thr Glu |     |
| 85 90 95  |     |
| AAT TTT AAC ATG TGG AAA AAT GAC ATG GTA GAA CAG ATG CAT GAG GAT | 336 |
| Asn Phe Asn Met Trp Lys Asn Asp Met Val Glu Gln Met His Glu Asp |     |
| 100 105 110   |     |
| ATA ATC AGT TTA TGG GAT CAA AGC CTA AAG CCA TGT GTA AAA TTA ACC | 384 |
| Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr |     |
| 115 120 125   |     |
| CCA CTC TGT GTT AGT TTA AAG TGC ACT GAT TTG GGG AAT GCT ACT AAT | 432 |
| Pro Leu Cys Val Ser Leu Lys Cys Thr Asp Leu Gly Asn Ala Thr Asn |     |
| 130 135 140   |     |
| ACC AAT AGT AGT AAT ACC AAT AGT AGT AGC GGG GAA ATG ATG ATG GAG | 480 |
| Thr Asn Ser Ser Asn Thr Asn Ser Ser Ser Gly Glu Met Met Met Glu |     |
| 145 150 155 160   |     |
| AAA GGA GAG ATA AAA AAC TGC TCT TTC AAT ATC AGC ACA AGC ATA AGA | 528 |
| Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn Ile Ser Thr Ser Ile Arg |     |

80

| 165   | 170 | 175 |      |
|---|-----|-----|------|
| GGT AAG GTG CAG AAA GAA TAT GCA TTT TTT TAT AAA CTT GAT ATA ATA<br>Gly Lys Val Gln Lys Glu Tyr Ala Phe Phe Tyr Lys Leu Asp Ile Ile<br>180 185 190     |     |     | 576  |
| CCA ATA GAT AAT GAT ACT ACC AGC TAT ACG TTG ACA AGT TGT AAC ACC<br>Pro Ile Asp Asn Asp Thr Thr Ser Tyr Thr Leu Thr Ser Cys Asn Thr<br>195 200 205     |     |     | 624  |
| TCA GTC ATT ACA CAG GCC TGT CCA AAG GTA TCC TTT GAG CCA ATT CCC<br>Ser Val Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro<br>210 215 220     |     |     | 672  |
| ATA CAT TAT TGT GCC CCG GCT GGT TTT GCG ATT CTA AAA TGT AAT AAT<br>Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asn<br>225 230 235 240 |     |     | 720  |
| AAG ACG TTC AAT GGA ACA GGA CCA TGT ACA AAT GTC AGC ACA GTA CAA<br>Lys Thr Phe Asn Gly Thr Gly Pro Cys Thr Asn Val Ser Thr Val Gln<br>245 250 255     |     |     | 768  |
| TGT ACA CAT GGA ATT AGG CCA GTA GTA TCA ACT CAA CTG CTG TTG AAT<br>Cys Thr His Gly Ile Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn<br>260 265 270     |     |     | 816  |
| GGC AGT CTA GCA GAA GAA GAG GTA GTA ATT AGA TCT GCC AAT TTC ACA<br>Gly Ser Leu Ala Glu Glu Glu Val Val Ile Arg Ser Ala Asn Phe Thr<br>275 280 285     |     |     | 864  |
| GAC AAT GCT AAA ACC ATA ATA GTA CAG CTG AAC CAA TCT GTA GAA ATT<br>Asp Asn Ala Lys Thr Ile Ile Val Gln Leu Asn Gln Ser Val Glu Ile<br>290 295 300     |     |     | 912  |
| AAT TGT ACA AGA CCC AAC AAC AAT ACA AGA AAA AGT ATC CGT ATC CAG<br>Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gln<br>305 310 315 320 |     |     | 960  |
| AGG GGA CCA GGG AGA GCA TTT GTT ACA ATA GGA AAA ATA GGA AAT ATG<br>Arg Gly Pro Gly Arg Ala Phe Val Thr Ile Gly Lys Ile Gly Asn Met<br>325 330 335     |     |     | 1008 |
| AGA CAA GCA CAT TGT AAC ATT AGT AGA GCA AAA TGG AAT GCC ACT TTA<br>Arg Gln Ala His Cys Asn Ile Ser Arg Ala Lys Trp Asn Ala Thr Leu<br>340 345 350     |     |     | 1056 |
| AAA CAG ATA GCT AGC AAA TTA AGA GAA CAA TTT GGA AAT AAT AAA ACA<br>Lys Gln Ile Ala Ser Lys Leu Arg Glu Gln Phe Gly Asn Asn Lys Thr<br>355 360 365     |     |     | 1104 |
| ATA ATC TTT AAG CAA TCC TCA GGA GGG GAC CCA GAA ATT GTA ACG CAC<br>Ile Ile Phe Lys Gln Ser Ser Gly Gly Asp Pro Glu Ile Val Thr His<br>370 375 380     |     |     | 1152 |
| AGT TTT AAT TGT GGA GGG GAA TTT TTC TAC TGT AAT TCA ACA CAA CTG<br>Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn Ser Thr Gln Leu<br>385 390 395 400 |     |     | 1200 |
| TTT AAT AGT ACT TGG TTT AAT AGT ACT TGG AGT ACT GAA GGG TCA AAT<br>Phe Asn Ser Thr Trp Phe Asn Ser Thr Trp Ser Thr Glu Gly Ser Asn<br>405 410 415     |     |     | 1248 |
| AAC ACT GAA GGA AGT GAC ACA ATC ACA CTC CCA TGC AGA ATA AAA CAA<br>Asn Thr Glu Gly Ser Asp Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln<br>420 425 430     |     |     | 1296 |
| TTT ATA AAC ATG GTG CAG GAA GTA GGA AAA GCA ATG TAT GCC CCT CCC<br>Phe Ile Asn Met Val Gln Glu Val Gly Lys Ala Met Tyr Ala Pro Pro<br>435 440 445     |     |     | 1344 |

SUBSTITUTE SHEET (RULE 26)

81

|   |      |
|---|------|
| ATC AGC GGA CAA ATT AGA TGT TCA TCA AAT ATT ACA GGG CTG CTA TTA<br>Ile Ser Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu Leu<br>450 455 460     | 1392 |
| ACA AGA GAT GGT GGT AAT AAC AAC AAT GGG TCC GAG ATC TTC AGA CCT<br>Thr Arg Asp Gly Gly Asn Asn Asn Asn Gly Ser Glu Ile Phe Arg Pro<br>465 470 475 480 | 1440 |
| GGA GGA GGA GAT ATG AGG GAC AAT TGG AGA AGT GAA TTA TAT AAA TAT<br>Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr<br>485 490 495     | 1488 |
| AAA GTA GTA AAA ATT GAA CCA TTA GGA GTA GCA CCC ACC AAG GCA AAG<br>Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys<br>500 505 510     | 1536 |
| AGA AGA GTG GTG CAG AGA GAA AAA TGA GCG G CCGC<br>Arg Arg Val Val Gln Arg Glu Lys<br>515 520  | 1571 |

## (2) INFORMATION FOR SEQ ID NO:26:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 522 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

|  |
|--|
| Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly<br>1 5 10 15       |
| Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg<br>20 25 30        |
| Gly Ala Arg Thr Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro<br>35 40 45        |
| Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys<br>50 55 60        |
| Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val<br>65 70 75 80     |
| Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Val Asn Val Thr Glu<br>85 90 95        |
| Asn Phe Asn Met Trp Lys Asn Asp Met Val Glu Gln Met His Glu Asp<br>100 105 110     |
| Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr<br>115 120 125     |
| Pro Leu Cys Val Ser Leu Lys Cys Thr Asp Leu Gly Asn Ala Thr Asn<br>130 135 140     |
| Thr Asn Ser Ser Asn Thr Asn Ser Ser Ser Gly Glu Met Met Met Glu<br>145 150 155 160 |
| Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn Ile Ser Thr Ser Ile Arg<br>165 170 175     |
| Gly Lys Val Gln Lys Glu Tyr Ala Phe Phe Tyr Lys Leu Asp Ile Ile<br>180 185 190     |
| Pro Ile Asp Asn Asp Thr Thr Ser Tyr Thr Leu Thr Ser Cys Asn Thr                    |



82

| 195   | 200 | 205         |
|---|-----|-------------|
| Ser Val Ile Thr Gln Ala Cys Pro Lys Val Ser Phe Glu Pro Ile Pro |     |             |
| 210   | 215 | 220         |
| Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Asn Asn |     |             |
| 225   | 230 | 235 240     |
| Lys Thr Phe Asn Gly Thr Gly Pro Cys Thr Asn Val Ser Thr Val Gln |     |             |
|   | 245 | 250 255     |
| Cys Thr His Gly Ile Arg Pro Val Val Ser Thr Gln Leu Leu Leu Asn |     |             |
|   | 260 | 265 270     |
| Gly Ser Leu Ala Glu Glu Glu Val Val Ile Arg Ser Ala Asn Phe Thr |     |             |
|   | 275 | 280 285     |
| Asp Asn Ala Lys Thr Ile Ile Val Gln Leu Asn Gln Ser Val Glu Ile |     |             |
|   | 290 | 295 300     |
| Asn Cys Thr Arg Pro Asn Asn Asn Thr Arg Lys Ser Ile Arg Ile Gln |     |             |
|   | 305 | 310 315 320 |
| Arg Gly Pro Gly Arg Ala Phe Val Thr Ile Gly Lys Ile Gly Asn Met |     |             |
|   | 325 | 330 335     |
| Arg Gln Ala His Cys Asn Ile Ser Arg Ala Lys Trp Asn Ala Thr Leu |     |             |
|   | 340 | 345 350     |
| Lys Gln Ile Ala Ser Lys Leu Arg Glu Gln Phe Gly Asn Asn Lys Thr |     |             |
|   | 355 | 360 365     |
| Ile Ile Phe Lys Gln Ser Ser Gly Gly Asp Pro Glu Ile Val Thr His |     |             |
|   | 370 | 375 380     |
| Ser Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn Ser Thr Gln Leu |     |             |
|   | 385 | 390 395 400 |
| Phe Asn Ser Thr Trp Phe Asn Ser Thr Trp Ser Thr Glu Gly Ser Asn |     |             |
|   | 405 | 410 415     |
| Asn Thr Glu Gly Ser Asp Thr Ile Thr Leu Pro Cys Arg Ile Lys Gln |     |             |
|   | 420 | 425 430     |
| Phe Ile Asn Met Val Gln Glu Val Gly Lys Ala Met Tyr Ala Pro Pro |     |             |
|   | 435 | 440 445     |
| Ile Ser Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly Leu Leu Leu |     |             |
|   | 450 | 455 460     |
| Thr Arg Asp Gly Gly Asn Asn Asn Asn Gly Ser Glu Ile Phe Arg Pro |     |             |
|   | 465 | 470 475 480 |
| Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu Tyr Lys Tyr |     |             |
|   | 485 | 490 495     |
| Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr Lys Ala Lys |     |             |
|   | 500 | 505 510     |
| Arg Arg Val Val Gln Arg Glu Lys                                 |     |             |
|   | 515 | 520         |

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1532 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..1522

(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

|   |     |
|---|-----|
| ATG GAT GCA ATG AAG AGA GGG CTC TGC TGT GTG CTG CTG CTG TGT GGA | 48  |
| Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly |     |
| 1 5 10 15   |     |
| GCA GTC TTC GTT TCG CCC AGC CAG GAA ATC CAT GCC CGA TTC AGA AGA | 96  |
| Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg |     |
| 20 25 30  |     |
| GGC GGC AGA GTA GAA AAG TTG TGG GTC ACA GTC TAT TAT GGG GTA CCT | 144 |
| Gly Gly Arg Val Glu Lys Leu Trp Val Thr Val Tyr Gly Val Pro     |     |
| 35 40 45  |     |
| GTG TGG AAA GAA GCA ACC ACC ACT CTA TTT TGT GCA TCA GAT GCT AAA | 192 |
| Val Trp Lys Glu Ala Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys     |     |
| 50 55 60  |     |
| GCA TAT GAT ACA GAG GTA CAT AAT GTT TGG GCC ACA CAT GCC TGT GTA | 240 |
| Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val |     |
| 65 70 75 80   |     |
| CCC ACA GAC CCC AAC CCA CAA GAA GTA GTA TTG GAA AAT GTA ACA GAA | 288 |
| Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Glu Asn Val Thr Glu |     |
| 85 90 95  |     |
| CAT TTT AAC ATG TGG AAA AAT AAC ATG GTA GAA CAG ATG CAG GAG GAT | 336 |
| His Phe Asn Met Trp Lys Asn Asn Met Val Glu Gln Met Gln Glu Asp |     |
| 100 105 110   |     |
| ATA ATC AGT TTA TGG GAT CAA AGC CTA AAG CCA TGT GTA AAA TTA ACC | 384 |
| Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr |     |
| 115 120 125   |     |
| CCA CTC TGT GTT ACT TTA AAT TGC AAG GAT GTG AAT GCT ACT AAT ACC | 432 |
| Pro Leu Cys Val Thr Leu Asn Cys Lys Asp Val Asn Ala Thr Asn Thr |     |
| 130 135 140   |     |
| ACT AAT GAT AGC GAG GGA ACG ATG GAG AGA GGA GAA ATA AAA AAC TGC | 480 |
| Thr Asn Asp Ser Glu Gly Thr Met Glu Arg Gly Glu Ile Lys Asn Cys |     |
| 145 150 155 160   |     |
| TCT TTC AAT ATC ACC ACA AGC ATA AGA GAT GAG GTG CAG AAA GAA TAT | 528 |
| Ser Phe Asn Ile Thr Thr Ser Ile Arg Asp Glu Val Gln Lys Glu Tyr |     |
| 165 170 175   |     |
| GCT CTT TTT TAT AAA CTT GAT GTA GTA CCA ATA GAT AAT AAT AAT ACC | 576 |
| Ala Leu Phe Tyr Lys Leu Asp Val Val Pro Ile Asp Asn Asn Asn Thr |     |
| 180 185 190   |     |
| AGC TAT AGG TTG ATA AGT TGT GAC ACC TCA GTC ATT ACA CAG GCC TGT | 624 |
| Ser Tyr Arg Leu Ile Ser Cys Asp Thr Ser Val Ile Thr Gln Ala Cys |     |
| 195 200 205   |     |
| CCA AAG ATA TCC TTT GAG CCA ATT CCC ATA CAT TAT TGT GCC CCG GCT | 672 |
| Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala |     |
| 210 215 220   |     |
| GGT TTT GCG ATT CTA AAG TGT AAT GAT AAG ACG TTC AAT GGA AAA GGA | 720 |
| Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Thr Phe Asn Gly Lys Gly |     |
| 225 230 235 240   |     |

84

|   |      |
|---|------|
| CCA TGT AAA AAT GTC AGC ACA GTA CAA TGT ACA CAT GGA ATT AGG CCA<br>Pro Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro<br>245 250 255     | 768  |
| GTA GTA TCA ACT CAA CTG CTG CTA AAT GGC AGT CTA GCA GAA GAA GAG<br>Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu<br>260 265 270     | 816  |
| GTA GTA ATT AGA TCT GAC AAT TTC ACG AAC AAT GCT AAA ACC ATA ATA<br>Val Val Ile Arg Ser Asp Asn Phe Thr Asn Asn Ala Lys Thr Ile Ile<br>275 280 285     | 864  |
| GTA CAG CTG AAA GAA TCT GTA GAA ATT AAT TGT ACA AGA CCC AAC AAC<br>Val Gln Leu Lys Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn<br>290 295 300     | 912  |
| AAT ACA AGA AAA AGT ATA CAT ATA GGA CCA GGG AGA GCA TTT TAT ACT<br>Asn Thr Arg Lys Ser Ile His Ile Gly Pro Gly Arg Ala Phe Tyr Thr<br>305 310 315 320 | 960  |
| ACA GGA GAA ATA ATA GGA GAT ATA AGA CAA GCA CAT TGT AAC ATT AGT<br>Thr Gly Glu Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Ile Ser<br>325 330 335     | 1008 |
| AGA GCA AAA TGG AAT GAC ACT TTA AAA CAG ATA GTT ATA AAA TTA AGA<br>Arg Ala Lys Trp Asn Asp Thr Leu Lys Gln Ile Val Ile Lys Leu Arg<br>340 345 350     | 1056 |
| GAA CAA TTT GAG AAT AAA ACA ATA GTC TTT AAT CAC TCC TCA GGA GGG<br>Glu Gln Phe Glu Asn Lys Thr Ile Val Phe Asn His Ser Ser Gly Gly<br>355 360 365     | 1104 |
| GAC CCA GAA ATT GTA ATG CAC AGT TTT AAT TGT GGA GGA GAA TTT TTC<br>Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu Phe Phe<br>370 375 380     | 1152 |
| TAC TGT AAT TCA ACA CAA CTG TTT AAT AGT ACT TGG AAT AAT AAT ACT<br>Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp Asn Asn Asn Thr<br>385 390 395 400 | 1200 |
| GAA GGG TCA AAT AAC ACT GAA GGA AAT ACT ATC ACA CTC CCA TGC AGA<br>Glu Gly Ser Asn Asn Thr Glu Gly Asn Thr Ile Thr Leu Pro Cys Arg<br>405 410 415     | 1248 |
| ATA AAA CAA ATT ATA AAC ATG GTG CAG GAA GTA GGA AAA GCA ATG TAT<br>Ile Lys Gln Ile Ile Asn Met Val Gln Glu Val Gly Lys Ala Met Tyr<br>420 425 430     | 1296 |
| GCC CCT CCC ATC AGA GGA CAA ATT AGA TGT TCA TCA AAT ATT ACA GGG<br>Ala Pro Pro Ile Arg Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly<br>435 440 445     | 1344 |
| CTG CTA TTA ACA AGA GAT GGT GGT ATT AAT GAG AAT GGG ACC GAG ATC<br>Leu Leu Leu Thr Arg Asp Gly Gly Ile Asn Glu Asn Gly Thr Glu Ile<br>450 455 460     | 1392 |
| TTC AGA CCT GGA GGA GGA GAT ATG AGG GAC AAT TGG AGA AGT GAA TTA<br>Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu<br>465 470 475 480 | 1440 |
| TAT AAA TAT AAA GTA GTA AAA ATT GAA CCA TTA GGA GTA GCA CCC ACC<br>Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr<br>485 490 495     | 1488 |
| AAG GCA AAG AGA AGA GTG GTG CAA AGA GAA AAA T GAGCGGCGCG<br>Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys<br>500 505                                    | 1532 |

## (2) INFORMATION FOR SEQ ID NO:28:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 507 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

```

Met Asp Ala Met Lys Arg Gly Leu Cys Cys Val Leu Leu Leu Cys Gly
 1           5           10           15
Ala Val Phe Val Ser Pro Ser Gln Glu Ile His Ala Arg Phe Arg Arg
 20           25           30
Gly Gly Arg Val Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro
 35           40           45
Val Trp Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys
 50           55           60
Ala Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val
 65           70           75           80
Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Glu Asn Val Thr Glu
 85           90           95
His Phe Asn Met Trp Lys Asn Asn Met Val Glu Gln Met Gln Glu Asp
100           105           110
Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val Lys Leu Thr
115           120           125
Pro Leu Cys Val Thr Leu Asn Cys Lys Asp Val Asn Ala Thr Asn Thr
130           135           140
Thr Asn Asp Ser Glu Gly Thr Met Glu Arg Gly Glu Ile Lys Asn Cys
145           150           155           160
Ser Phe Asn Ile Thr Thr Ser Ile Arg Asp Glu Val Gln Lys Glu Tyr
165           170           175
Ala Leu Phe Tyr Lys Leu Asp Val Val Pro Ile Asp Asn Asn Asn Thr
180           185           190
Ser Tyr Arg Leu Ile Ser Cys Asp Thr Ser Val Ile Thr Gln Ala Cys
195           200           205
Pro Lys Ile Ser Phe Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala
210           215           220
Gly Phe Ala Ile Leu Lys Cys Asn Asp Lys Thr Phe Asn Gly Lys Gly
225           230           235           240
Pro Cys Lys Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro
245           250           255
Val Val Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu
260           265           270
Val Val Ile Arg Ser Asp Asn Phe Thr Asn Asn Ala Lys Thr Ile Ile
275           280           285
Val Gln Leu Lys Glu Ser Val Glu Ile Asn Cys Thr Arg Pro Asn Asn
290           295           300
Asn Thr Arg Lys Ser Ile His Ile Gly Pro Gly Arg Ala Phe Tyr Thr

```

86

```

305          310          315          320
Thr Gly Glu Ile Ile Gly Asp Ile Arg Gln Ala His Cys Asn Ile Ser
      325          330          335
Arg Ala Lys Trp Asn Asp Thr Leu Lys Gln Ile Val Ile Lys Leu Arg
      340          345          350
Glu Gln Phe Glu Asn Lys Thr Ile Val Phe Asn His Ser Ser Gly Gly
      355          360          365
Asp Pro Glu Ile Val Met His Ser Phe Asn Cys Gly Gly Glu Phe Phe
      370          375          380
Tyr Cys Asn Ser Thr Gln Leu Phe Asn Ser Thr Trp Asn Asn Asn Thr
      385          390          395          400
Glu Gly Ser Asn Asn Thr Glu Gly Asn Thr Ile Thr Leu Pro Cys Arg
      405          410          415
Ile Lys Gln Ile Ile Asn Met Val Gln Glu Val Gly Lys Ala Met Tyr
      420          425          430
Ala Pro Pro Ile Arg Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly
      435          440          445
Leu Leu Leu Thr Arg Asp Gly Gly Ile Asn Glu Asn Gly Thr Glu Ile
      450          455          460
Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg Ser Glu Leu
      465          470          475          480
Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly Val Ala Pro Thr
      485          490          495
Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys
      500          505

```

## (2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```

Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg
1          5          10          15

```

What is claimed is:

1. A recombinant nucleic acid molecule which encodes a mutant HIV-1 gp120 envelope glycoprotein comprising a V3 loop deletion and a C4 domain<sub>(W->X)</sub> point mutation, wherein X is an amino acid residue other than tryptophan.
2. The recombinant nucleic acid molecule of claim 1, wherein X is a valine residue.
3. The recombinant nucleic acid molecule of claim 1, wherein the nucleic acid molecule is a DNA molecule.
4. The recombinant nucleic acid molecule of claim 3, wherein the DNA molecule is a plasmid.
5. The recombinant nucleic acid molecule of claim 4, wherein the plasmid comprises the sequence of the plasmid designated PPI4-tPA.
6. The recombinant nucleic acid molecule of claim 1, wherein the C4 domain is an HIV-1<sub>LA1</sub> gp120 envelope glycoprotein C4 domain.
7. The recombinant nucleic acid molecule of claim 6, wherein the mutant HIV-1 gp120 envelope glycoprotein is a mutant HIV-1<sub>LA1</sub> gp120 envelope glycoprotein.
8. The recombinant nucleic acid molecule of claim 1, wherein the C4 domain is an HIV-1<sub>JR-FL</sub> gp120 envelope glycoprotein C4 domain.
9. The recombinant nucleic acid molecule of claim 8,

wherein the mutant HIV-1 gp120 envelope glycoprotein is a mutant HIV-1<sub>JR-FL</sub> gp120 envelope glycoprotein.

10. The mutant HIV-1 gp120 envelope glycoprotein encoded by the recombinant nucleic acid molecule of claim 1.
11. A vaccine which comprises a therapeutically effective amount of the mutant HIV-1 gp120 envelope glycoprotein of claim 10, and an adjuvant.
12. A method of treating an HIV-1-infected subject, which comprises immunizing the HIV-1-infected subject with the vaccine of claim 11, thereby treating the HIV-1-infected subject.
13. A vaccine which comprises a prophylactically effective amount of the mutant HIV-1 gp120 envelope glycoprotein of claim 10, and an adjuvant.
14. A method of reducing the likelihood of an HIV-1-exposed subject's becoming infected with HIV-1, which comprises immunizing the HIV-1-exposed subject with the vaccine of claim 13, thereby reducing the likelihood of the HIV-1-exposed subject's becoming infected with HIV-1.
15. A method of reducing the likelihood of a non-HIV-1-exposed subject's becoming infected with HIV-1, which comprises immunizing the non-HIV-1-exposed subject with the vaccine of claim 13, thereby reducing the likelihood of the non-HIV-1-exposed subject's becoming infected with HIV-1.
16. A method of obtaining partially purified antibodies which specifically bind to the CD4-binding domain of

HIV-1 gp120 envelope glycoprotein, which method comprises (a) immunizing a non-HIV-1-exposed subject with the vaccine of claim 13, (b) recovering from the immunized subject serum comprising said antibodies, and  
5 (c) partially purifying said antibodies, thereby obtaining partially purified antibodies which specifically bind to the CD4-binding domain of HIV-1 gp120 envelope glycoprotein.

10 17. The method of claim 16, wherein the subject is a human.

18. The partially purified antibodies produced by the method of claim 16.

15 19. A pharmaceutical composition, which comprises a therapeutically effective amount of the partially purified antibodies of claim 18, and a pharmaceutically acceptable carrier.

20 20. A method of treating an HIV-1-infected subject, which comprises administering to the subject a dose of the pharmaceutical composition of claim 19 effective to reduce the population of HIV-1-infected cells in the HIV-1-infected subject, thereby treating the HIV-1-  
25 infected subject.

21. A method of treating an HIV-1-infected subject, which comprises administering to the subject a dose of the pharmaceutical composition of claim 19 effective to  
30 reduce the population of HIV-1 in the HIV-1-infected subject, thereby treating the HIV-1-infected subject.

22. A composition which comprises a prophylactically effective amount of the partially purified antibodies  
35 of claim 18, and a pharmaceutically acceptable carrier.



23. A method of reducing the likelihood of an HIV-1-exposed subject's becoming infected with HIV-1, which comprises administering to the HIV-1-exposed subject a dose of the composition of claim 22 effective to reduce the population of HIV-1 in the HIV-1-exposed subject, thereby reducing the likelihood of the subject's becoming infected with HIV-1.

24. The method of claim 23, wherein the subject is a medical practitioner.

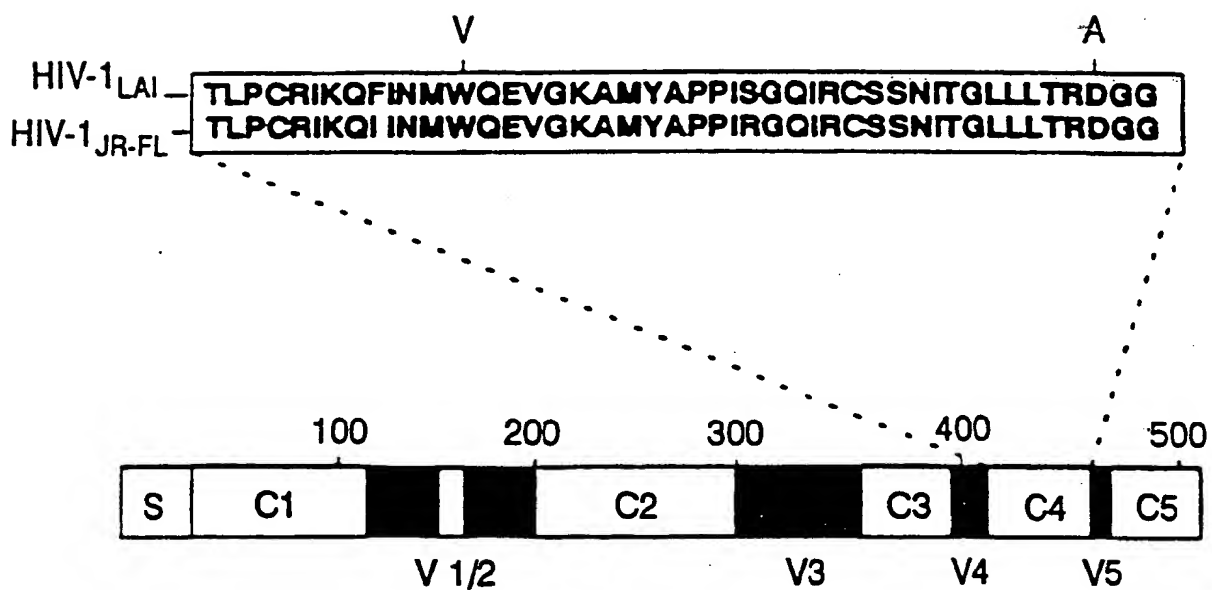
25. The method of claim 23, wherein the subject is a newborn infant.

26. A method of reducing the likelihood of a non-HIV-1-exposed subject's becoming infected with HIV-1 as a result of exposure thereto during an incident wherein there is an increased risk of exposure to HIV-1, which comprises administering to the subject immediately prior to the incident a dose of the composition of claim 22 effective to reduce the population of HIV-1 to which the subject is exposed during the incident, thereby reducing the likelihood of the subject's becoming infected with HIV-1.

27. The method of claim 26, wherein the subject is a medical practitioner.

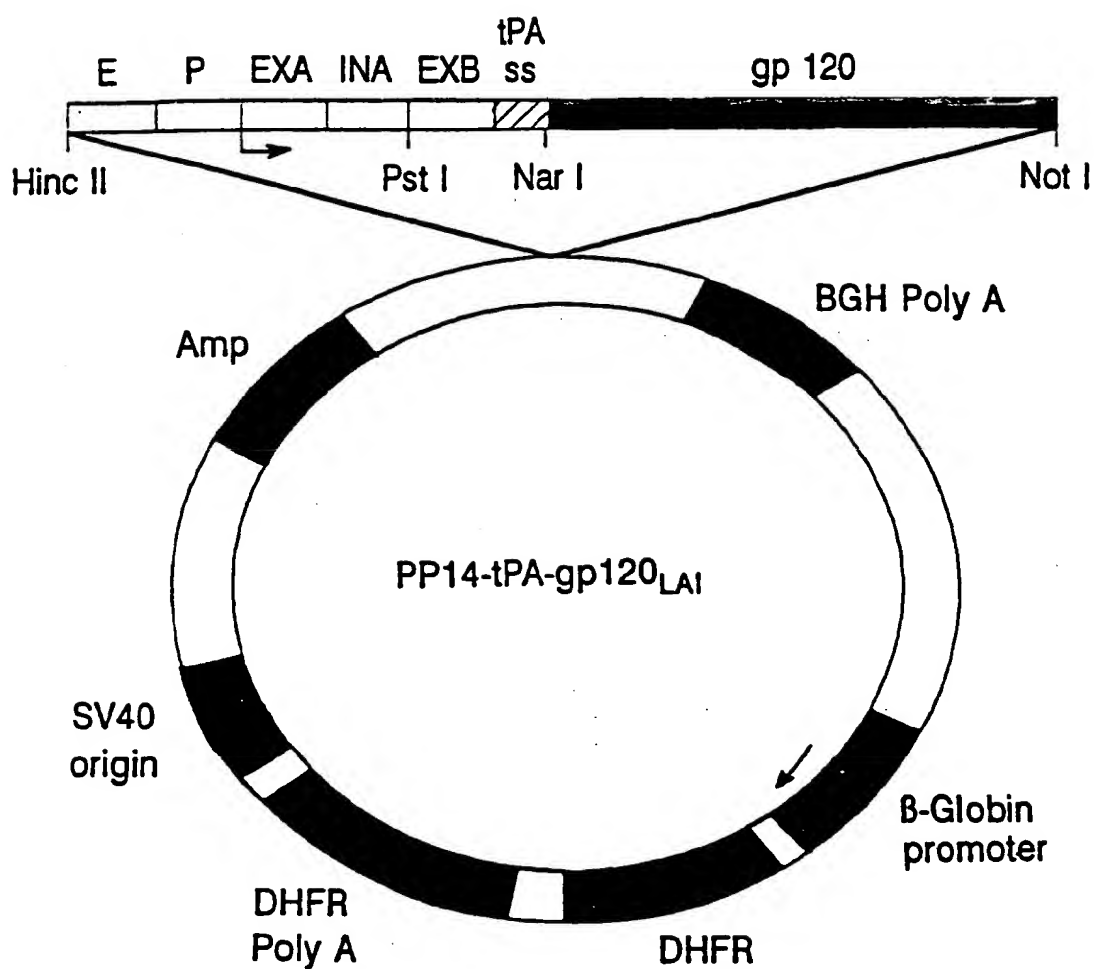
1/42

FIGURE 1



2/42

FIGURE 2



3/42

|           |
|-----------|
| FIGURE 3A |
| FIGURE 3B |
| FIGURE 3C |
| FIGURE 3D |
| FIGURE 3E |
| FIGURE 3F |

## FIGURE 3A

HincII

```

1  ttgacattgattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatgga
73  gttccgcgttacataacttacggtaatggccgcctggctgacggcccaacgaccccccattgacgtc
145  aataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggactatttacg
217  gtactgcccacttggcagtagacatcaagtgtatcatatgccaaagtacgccccctattgacgtcaatgacgg
289  taaatggccgcctggcattatgcccagtagaccttatgggactttcctacttggcagtacatctacgt
361  attagtcacgcctattaccatggtgatgcggttttggcagtagacatcaatgggcgtggatagcgggttggactc
433  acggggatttccaagtctccaccccatgacgtcaatgggagtttgttttggcaccacaaatcaacggggactt

```

**FIGURE 3B**

505 tccaaaatgtcgtaacaaactccgccccattgacgcaaatgggcggtaggcgtgtacgggtgggaggtctatat  
Exon A  
577 aagcagagctcgtttagtgaaacgctcagatgccctggagacgcccatccacgctgtttttgacctccatagaag  
→ Transcription Start  
649 ACACGGGACCGATCCAGCCTCCGGCGCGGACGGTGCAATTGGAACGGGATTCGCCGTGCCAAGAGTGA  
Intron A  
721 Cgtaagtaccgcctatagactctataggcacacacccctttggctcttatgcatgctatactgtttttggcttg  
793 ggccaaacacccgcctcctagataggtgatggtatagcttagcctataggtgtgggttattgaccattatttgac  
865 cactccctattgggtgacgatactttccattactaatccataaacatggccgctctttgccacaaactatctct  
937 attggctatatgccaatactctgtccttcagagactgacacggactctgtatttttacaggatgggggtccca  
1009 ttattattacaaaattcacatatatacaaacgcccgtccccgtgcccgcagtttttattaacatgcgggat  
1081 ctccacgcgaatctcgggtacgtgttcgggacatgggctcttctccggtagcggcgagctccaacatccgag  
1153 cctgtcccatgcccatgcctccagcggctcatggtcgctcggcagctccttgctcctaacagtgagggccag  
1225 acttaggcacaggacaatgccaccaccacagtggtgccgcacaaggccgtggcggtagggtatgtgtctga

5/42

## FIGURE 3C

```

1297 aaatgagctcggagattgggctcgaccgctgacgcagatggaagacttaaggcagcgagaaagatgc
1369 aggcagctgagttgttattctgtagagttggaggttaactccggttgcggtgctgttaacggtggagggca

1441 gtgtagtctgagcagtaactcgttgctgcgcgcgcgcaccagacataatagctgacagactaacagactgt
      PstI      Exon B      tPA signal sequence
1513 tcctttccatgggtcttttctgcagTCACCGTCCTTGACACGATGGATGCAATGAAGAGAGGGCTCTGCTGT
      1          M D A M K R G L C C
1585 GTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCGCCAGGAAATCCCATGCCCGATTTCAGAAGAGCGGCC
      11         V L L L C G A V F V S P S Q E I H A R F R R G A
1657 AGAACAGAAAAATTGTGGGTCACAGTCTATTATGGGTACCTGTGTGGAAGGAAGCAACCACCTCTATTT
      35         R T E K L W V T V Y Y G V P V W X E A T T T L F
      ▲ Signal cleavage

1729 TGTGCATCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCCACA
      59         C A S D A K A Y D T E V H N V W A T H A C V P T
1801 GACCCCAACCCACAAGAAGTAGTATGGTAAATGTGACAGAAAAATTTTAACATGTGGAAAAATGACATGGTA
      83         D P N P Q E V V L V N V T E N F N M W K N D M V

```

6/42

## FIGURE 3D

```

1873 GAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAGCCCTAAAGCCATGTGTAAATTAACCCCACTC
107   E Q M H E D I I S L W D Q S L K P C V K L T P L

1945 TGTGTTAGTTTAAAGTGCACTGATTGGGGAATGCTACTAATACCAATAGTAGTAATACCAATAGTAGTAGC
131   C V S L K C T D L G N A T N T N S S N T N S S S

2017 GGGGAAATGATGATGGAGAAAGGAGAGATAAAACTGCTCTTTCAATATCAGCACAAAGCATAAGAGGTAAG
155   G E M M E K G E I K N C S F N I S T S I R G K

2089 GTGCAGAAAGAATATGCAATTTTTTATAAAGTTGATATAATACCAATAGATAATGATACTACCAGCTATACG
179   V Q K E Y A F F Y K L D I I P I D N D T T S Y T

2161 TTGACAAGTTGTAACACCTCAGTCATTACACAGGCGCTGTCCAAGGTATCCTTTGAGCCCAATTCCCATACAT
203   L T S C N T S V I T Q A C P K V S F E P I P I H

2233 TATTGTGCCCCGGCTGTTTTCGGATTCTAAATGTAATAAAGACGTTCAATGGAACAGGACCATGTACA
227   Y C A P A G F A I L K C N N K T F N G T G P C T

2305 AATGTCAGCACAGTACATGTACACATGGAATTAGGCCAGTAGTATCAACCTCAACTGCTGTTGAATGGCAGT
251   N V S T V Q C T H G I R P V V S T Q L L L N G S

2377 CTAGCAGAGAGAGGTAGTAATTAGATCTGCCAATTTACACAGACAATGCTAAACCATTAATAGTACAGCTG
275   L A E E E V V I R S A N F T D N A K T I I V Q L

```

7/42

## FIGURE 3E

2449 AACCAATCTGTAGAAATTAATTGTACAAGACCCCAACAACAATACAAGAAAAGTATCCGTATCCAGAGGGGA  
 299 N Q S V E I N C T R P N N N T R K S I R I Q R G  
  
 2521 CCAGGAGAGCATTTGTTACAATAGGAAAATAGGAAATATGAGACAAGCACATTGTAAACATTAGTAGAGCA  
 323 P G R A F V T I G K I G N M R Q A H C N I S R A  
  
 2593 AAATGGAATGCCACTTTTAAACAGATAGCTAGCAAAATTAAGAGAACAATTTGGAATAATAAACAATAATC  
 347 K W N A T L K Q I A S K L R E Q F G N N K T I I  
  
 2665 TTTAAGCAATCCTCAGGAGGGACCCAGAAATTGTAAACGCACAGTTTAAATTGTGGAGGGGAATTTTCTAC  
 371 F K Q S S G G D P E I V T H S F N C G G E F Y  
  
 2737 TGTAAATTCACACAACACTGTTAATAGTACTTGGTTAATAGTACTTGGAGTACTGAAGGGTCAATAACACT  
 395 C N S T Q L F N S T W F N S T W S T E G S N N T  
  
 2809 GAAGGAAGTGACACAATCACACTCCCATGCAGAAATAAACAATTTATAAACATGTGGCAGGAAGTAGGAAA  
 419 E G S D T I T L P C R I K Q F I N M W Q E V G K  
  
 2881 GCAATGTATGCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAAAATATTACAGGGCTGCTATTAACAAGA  
 443 A M Y A P P I S G Q I R C S S N I T G L L L T R  
  
 2953 GATGGTGTAAACAACAATGGGTCGAGATCTTCAGACCCTGGAGGAGGAGATATGAGGACAATTGGAGA  
 467 D G G N N N G S E I F R P G G G D M R D N W R



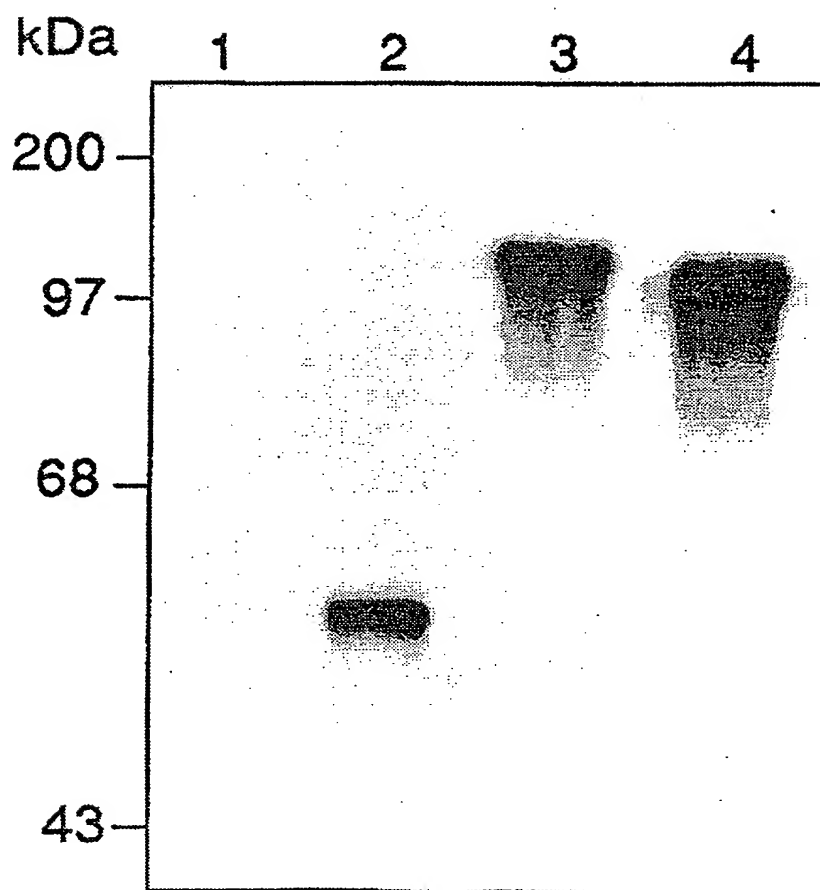
8/42

## FIGURE 3F

3025 AGTGAATTATATATAAAGTAGTAGTAAAAATTGAACCATTTAGGAGTAGCACCCACCAAGGCAAAGAGAAGA  
491 S E L Y K Y K V V K V I E P L G V A P T K A K R R  
NotI  
3097 GTGGTGCAGAGAGAAAAATGAGCGGCGCGC  
515 V V Q R E K -

9/42

FIGURE 4



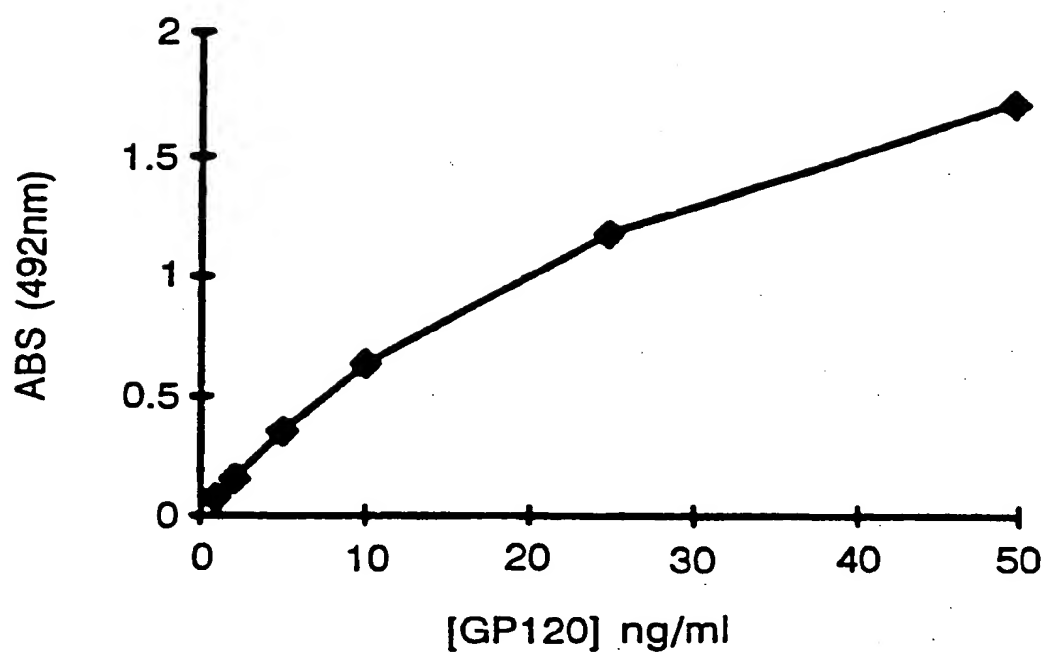
10/42

FIGURE 5A

| Stable CHO<br>clone | [gp120]<br>(ng/ml) |
|---------------------|--------------------|
| 5                   | 6                  |
| 6                   | 14                 |
| 9                   | 123                |
| 10                  | 4                  |
| 12                  | 18                 |
| 13                  | 18                 |

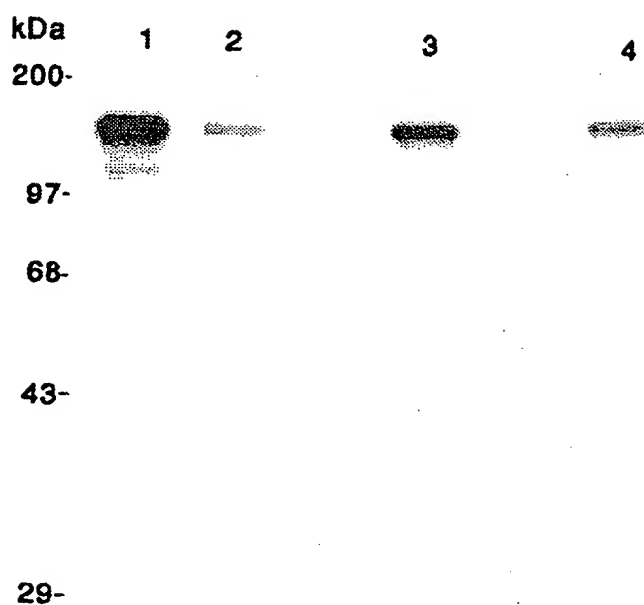
11/42

FIGURE 5B



12/42

FIGURE 6



13/42

## FIGURE 7A

|           |
|-----------|
| FIGURE 7A |
| FIGURE 7B |
| FIGURE 7C |

JR-FL

1 ATGGATGCAATGAAGAGA  
 1 M D A M K R  
  
 19 GGGCTCTGCTGTGCTGTGCTGTGGAGCAGTCTTCGTTTCGCCAGCCAGGAATC  
 7 G L C C V L L L C G A V F V S P S Q E I  
 NaI  
 79 CATGCCCGATTTCAGAAAGAGGGCGCAGAGTAGAAAGTTGTGGGTCACAGTCTATTATGGG  
 27 H A R F R R G A R V E K L W V T V Y Y G  
 ▲ Signal cleavage  
 139 GTACCTGTGTGGAAGAAGCAACCACCCTCTATTTTGTGCATCAGATGCTAAAGCATAT  
 47 V P V W K E A T T T L F C A S D A K A Y  
  
 199 GATACAGAGGTACATAATGTTTGGGCCACACATGCCCTGTGTACCCACAGACCCCAACCCA  
 67 D T E V H N V W A T H A C V P T D P N P  
  
 259 CAAGAAGTAGTATTGGAAATGTAACAGAACATTTTAACATGTGGAATAAATACATGGTA  
 87 Q E V V L E N V T E H F N M W K N N M V  
  
 319 GAACAGATGCAGGAGGATATAATCAGTTTATGGGATCAAAGCCTAAAGCCATGTGTAAAA  
 107 E Q M Q E D I I S L W D Q S L K P C V K  
  
 379 TTAACCCCACTCTGTGTACTTTAAATTGCAAGGATGTGAATGCTACTAATACCACCTAAT  
 127 L T P L C V T L N C K D V N A T N T T N  
  
 439 GATAGCGGGAACGATGGAGAGAGGAGAAATAAAACTGCTCTTTCAATATCACCACA  
 147 D S E G T M E R G E I K N C S F N I T T

14/42

## FIGURE 7B

499 AGCATAAGAGATGAGGTGCAGAAAGAATATGCTCTTTTATATAAAGTGTAGTACCA  
 167 S I R D E V Q K E Y A L F Y K L D V V P  
  
 559 ATAGATAATAATAACCAGCTATAGTTGATAAGTTGTGACACCTCAGTCATTACACAG  
 187 I D N N N T S Y R L I S C D T S V I T Q  
  
 619 GCCTGTCCAAGATATCCTTTGAGCCAATTCCCATACATTATTGTGCCCGGCTGTTT  
 207 A C P K I S F E P I P I H Y C A P A G F  
  
 679 GCGATTCTAAGTGTAATGATAAGACGTTCAATGGAAAGGACCATGTAAAAATGTCAGC  
 227 A I L K C N D K T F N G K G P C K N V S  
  
 739 ACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGCTAAATGGC  
 247 T V Q C T H G I R P V V S T Q L L L N G  
  
 799 AGTCTAGCAGAAGAGGTAGTAATTAGATCTGACAATTTTCACGAACAATGCTAAACC  
 267 S L A E E V V I R S D N F T N N A K T  
  
 859 ATAATAGTACAGCTGAAGAATCTGTAGAAATTAATTGTACAAGACCCCAACAATACA  
 287 I I V Q L K E S V E I N C T R P N N T  
  
 919 AGAAAAGTATACATATAGGACCAGGAGAGCATTTTATCTACAGGAGAAATAATAGGA  
 307 R K S I H I G P G R A F Y T T G E I I G  
  
 979 GATATAAGACACGACATTTGTAACATTAGTAGAGCAAAATGGAATGACACTTTAAACAG  
 327 D I R Q A H C N I S R A K W N D T L K Q

15/42

## FIGURE 7C

1039 ATAGTTATAAAATTAGAGAACAAATTTGAGAATAAAACAATAGTCTTTAATCACTCCTCA  
 347 I V I K L R E Q F E N K T I V F N H S S  
  
 1099 GGAGGGACCCAGAAATTGTAATGCACAGTTTAAATTGGAGGAGAATTTTCTACTGT  
 367 G G D P E I V M H S F N C G G E F F Y C  
  
 1159 AATCAACACAACACTGTTTAATAGTACTTGGATAATAATACTGAAGGGTCAATAACACT  
 387 N S T Q L F N S T W N N N T E G S N N T  
  
 1219 GAAGGAAATACTATCACACTCCCATGCAGAAATAAAACAATTAATAACAATGTGCAGGAA  
 407 E G N T I T L P C R I K Q I I N M W Q E  
  
 1279 GTAGGAAAGCAATGTATGCCCTCCCATCAGAGGACAAATTAGATGTTTCATCAAAATAT  
 427 V G K A M Y A P P I R G Q I R C S S N I  
  
 1339 ACAGGGCTGCTATTAAACAAGAGATGGTGTATTAAATGAGAAATGGACCGAGATCTTCAGA  
 447 T G L L L T R D G G I N E N G T E I F R  
  
 1399 CCTGGAGGAGATATGAGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTA  
 467 P G G G D M R D N W R S E L Y K Y K V V  
  
 1459 AAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAGAGAGAGTGGTGCAAGAGAA  
 487 K I E P L G V A P T K A K R R V V Q R E  
  
 1519 AAATGAGGGGGCGC  
 507 K

NotI



FIGURE 8A

|           |
|-----------|
| FIGURE 8A |
| FIGURE 8B |
| FIGURE 8C |

LAI ΔV3

1 ATGGATGCAATGAAGAGAGGGCTCTGTGTGTGCTG  
1 M D A M K R G L C C V L  
37 CTGCTGTGTGGAGCAGTCTTCGTTTCGCCAGCCAGGAATCCATGCCGATTTCAGAAGAGGCGGCCAGAACAA  
13 L L C G A V F V S P S Q E I H A R F R R G A R T  
109 GAAAATTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCACCCTCTATTTTGTGCA  
37 E K L W V T V Y Y G V P V W K E A T T T L F C A  
181 TCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCCTGTGTACCCACAGACCCC  
61 S D A K A Y D T E V H N V W A T H A C V P T D P  
253 AACCCACAAGTAGTATTGGTAAATGTGACAGAAATTTTAACATGTGGAAATGACATGGTAGAACAG  
85 N P Q E V V L V N V T E N F N M W K N D M V E Q  
325 ATGCATGAGGATATAATCAGTTTATGGGATCAAGCCCTAAAGCCATGTGTAAATTAACCCCACTCTGTGT  
109 M H E D I I S L W D Q S L K P C V K L T P L C V  
397 AGTTTAAAGTGCACCTGATTGGGGAATGCTACTATAATACCAATAGTAGTAATAGTAGCGGGAA  
133 S L K C T D L G N A T N T N S S N T N S S S G E

NarI

Signal cleavage ▲

17/42

## FIGURE 8B

469 ATGATGATGGAGAAAGGAGAGATAAAAACCTGCTCTTTCAATATCAGCACAGCATAAGAGGTAAGGTGCAG  
157 M M M E K G E I K N C S F N I S T S I R G K V Q

541 AAAGAATATGCATTTTTTATAAACTTGATATAATACCAATAGATAATGATACTACCAGCTATACGTTGACA  
181 K E Y A F F Y K L D I I P I D N D T T S Y T L T

613 AGTTGTAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAGCCCAATTTCCCATACATTATTGT  
205 S C N T S V I T Q A C P K V S F E P I P I H Y C

685 GCCCCGGCTGTTTGGGATTCTAAAATGTAATAAGACGTTCAATGGAACAGGACCATGTACAAATGTC  
229 A P A G F A I L K C N N K T F N G T G P C T N V

757 AGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTGTAATGGCAGTCTAGCA  
253 S T V Q C T H G I R P V V S T Q L L L N G S L A

829 GAAGAAGAGGTAGTAATTAGATCTGCCAATTTACACAGACAATGCTAAACCAATAATAGTACAGCTGAACCAA  
277 E E E V V I R S A N F T D N A K T I I V Q L N Q

901 TCTGTAGAAATTAATTGTACAGGTGCTGGACATTGTAACATTAGTAGAGCAAAATGGAATGCCACTTTAAAA  
301 S V E I N C T G A G H C N I S R A K W N A T L K

18/42

## FIGURE 8C

973 CAGATAGCTAGCAAAATTAAGAGAACAATTTGGAAATAATAAAACAATAATCTTTAAGCAATCCTCAGGAGGG  
 325 Q I A S K L R E Q F G N N K T I I F K Q S S G G  
 1045 GACCCAGAAAATTGTAACGCACAGTTTAAATTGTGGAGGGGAATTTTCTACTGTAATTCAACACAACACTGTTT  
 349 D P E I V T H S F N C G G E F F Y C N S T Q L F  
 1117 AATAGTACTTGGTTTAAATAGTACTTGGAGTACTGAAGGGTCAATAAACAATAACACTGAAGGAAGTGACACAATCACA  
 373 N S T W F N S T W S T E G S N N T E G S D T I T  
 1189 CTCCCATGCAGAATAAAACAATTTATAAACAATGTGCAGGAAGTAGGAAAGCAATGTATGCCCTCCCATC  
 397 L P C R I K Q F I N M W Q E V G K A M Y A P P I  
 1261 AGCGGACAAATTAGATGTTCATCAATAATATTACAGGGCTGCTATTAAACAAGAGATGGTGGTAATAACAACAAT  
 421 S G Q I R C S S N I T G L L L T R D G G N N N  
 1333 GGGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGACAAATTGGAGAAGTGAATTATATAATATAAA  
 445 G S E I F R P G G G D M R D N W R S E L Y K Y K  
 1405 GTAGTAAAAATTGAACCATTAGGAGTAGCACCCCAAGCAAGAGAGAGTGGTGCAGAGAGAAAATGA  
 469 V V K I E P L G V A P T K A K R R V V Q R E K -  
 NotI  
 1477 GCGGCGCG



20/42

## FIGURE 9B

439 GATAGCGAGGGAACGATGGAGAGAGGAGAAATAAAAACTGCTCTTTCAATATCACCACA  
 147 D S E G T M E R G E I K N C S F N I T T  
  
 499 AGCATAAGAGATGAGGTGCAGAAAGAATAATGCTCTTTTATATAAAC TTGATGTAGTACCA  
 167 S I R D E V Q K E Y A L F Y K L D V V P  
  
 559 ATAGATAATAATAACCAGCTATAGGTTGATAAGTTGTGACACCTCAGTCATTACACAG  
 187 I D N N N T S Y R L I S C D T S V I T Q  
  
 619 GCCTGTCCAAAGATATCCTTTGAGCCAAATCCCATACATTATTGTGCCCGCTGGTTTT  
 207 A C P K I S F E P I P I H Y C A P A G F  
  
 679 GCGATTCTAAAGTGTAATGATAAGACGTTCAATGGAAAAGGACCATGTAAAAATGTCAGC  
 227 A I L K C N D K T F N G K G P C K N V S  
  
 739 ACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGCTAAATGGC  
 247 T V Q C T H G I R P V V S T Q L L L N G  
  
 799 AGTCTAGCAGAAGAGAGGTAGTAATTAGATCTGACAATTTACGAAACAATGCTAAACC  
 267 S L A E E E V V I R S D N F T N N A K T  
  
 859 ATAATAGTACAGCTGAAAGAATCTGTAGAAATTAATTGTACAGGCTGCGACATTGTAAAC  
 287 I I V Q L K E S V E I N C T G A G H C N  
  
 919 ATTAGTAGCAAAATGGAATGACACTTTAAACAGATAGTTATAAAATTAAGAGAACAA  
 307 I S R A K W N D T L K Q I V I K L R E Q

21/42

## FIGURE 9C

979 TTTGAGATAAAACAATAGTCTTTAATCACTCCTCAGGAGGGACCCAGAAATTGTAATG  
 327 F E N K T I V F N H S S G G D P E I V M  
  
 1039 CACAGTTTAAATTGTGGAGGAGAATTTTCTACTGTAAATTCACACAACTGTTTAATAGT  
 347 H S F N C G G E F F Y C N S T Q L F N S  
  
 1099 ACTTGGAAATAATACTGAAGGGTCAAAATAACACTGAAGGAAATACATCACACTCCCA  
 367 T W N N N T E G S N N T E G N T I T L P  
  
 1159 TGCAGAAATAAAACAATTATAAACATGTGGCAGGAAGTAGGAAAGCAATGTATGCCCT  
 387 C R I K Q I I N M W Q E V G K A M Y A P  
  
 1219 CCCATCAGAGGACAAATTAGATGTTTCATCAATAATTACAGGGCTGCTATTAAACAAGAGAT  
 407 P I R G Q I R C S S N I T G L L T R D  
  
 1279 GGTGGTATTATGAGAATGGACCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGAC  
 427 G G I N E N G T E I F R P G G G D M R D  
  
 1339 AATTGGAGAAGTGAATTATATAATAATAAGTAGTAAATAATTGAACCATTAGGAGTAGCA  
 447 N W R S E L Y K Y K V V K I E P L G V A  
  
 1399 CCCACCAAGGCAAGAGAAGAGTGGTGCAAGAGAGAAATGAGGGGCGC  
 487 P T K A K R R V V Q R E K -

Noti

22/42

## FIGURE 10A

|            |
|------------|
| FIGURE 10A |
| FIGURE 10B |
| FIGURE 10C |

LAI ΔV3-CD4<sup>-</sup>

1 ATGGATGCAATGAAGAGAGGGCTCTGTGTGTGCTG  
 1 M D A M K R G L C C V L  
 37 CTGCTGTGGAGCAGTCTTGGTTTCGCCCCAGCCAGGAATCCATGCCCGATTTCAGAAGAGGGCGCCAGAACA  
 13 L L C G A V F V S P S Q E I H A R F R R G A R T  
 109 GAAAAATTGTGGTCACAGTCTATTATGGGGTACCTGTGTGGAAGGAAGCAACCACCTCTATTTTGTGCA  
 37 E K L W V T V Y Y G V P V W K E A T T T L F C A  
 181 TCAGATGCTAAAGCATATGATACAGAGGTACATAATGTTTGGGCCACACATGCCTGTGTACCCACAGACCCC  
 61 S D A K A Y D T E V H N V W A T H A C V P T D P  
 253 AACCACAAGAAGTAGTATTGGTAAATGTGACAGAAAATTTAACATGTGGAATAATGACATGTGTAGAACAG  
 85 N P Q E V V L V N V T E N F N M W K N D M V E Q  
 325 ATGCATGAGGATATAATCAGTTTATGGGATCAAGCCCTAAGCCATGTGTAAATAATTAACCCCACTCTGTGTT  
 109 M H E D I I S L W D Q S L K P C V K L T P L G V  
 397 AGTTAAAGTGCACCTGATTGGGGAATGCTACTAATAACCAATAGTAGTAATACCAATAGTAGCGGGAA  
 133 S L K C T D L G N A T N T N S S N T N S S S G E

Signal cleavage ▲

23/42

## FIGURE 10B

```

469 ATGATGATGGAGAAAGGAGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAAGCATAAGAGGTAAGGTGCAG
157 M M E K G E I K N C S F N I S T S I R G K V Q

541 AAAGAATATGCATTTTTTATAACTTGATATAATACCAATAGATAATGATACCTACCAGCTATACGTTGACA
181 K E Y A F F Y K L D I I P I D N D T T S Y T L T

613 AGTTGTAACACCTCAGTCATTACACAGGCCCTGTCCAAAGGTATCCTTTGAGCCCAATTCCCATACATTATTGT
205 S C N T S V I T Q A C P K V S F E P I P I H Y C

685 GCCCCGGCTGGTTTTCGATTCTAAANTGTAATAAAGACGTTCAATGGAACAGGACCATGTACAAATGTC
229 A P A G F A I L K C N N K T F N G T G P C T N V

757 AGCACAGTACAAATGTACACATGGAATTAGGCCAGTAGTATCAACTCAACTGCTGTTGAATGGCAGTCTAGCA
253 S T V Q C T H G I R P V V S T Q L L L N G S L A

829 GAAGAAGAGGTAGTAATTAGATCTGCCAATTTACAGACAAATGCTAAACCAATAATAGTACAGCTGAACCAA
277 E E E V V I R S A N F T D N A K T I I V Q L N Q

901 TCTGTAGAAATTAATTGTACAGGTGCTGGACATTGTAAACATTAGTAGAGCAAAATGGAATGCCACTTTAAAA
301 S V E I N C T G A G H C N I S R A K W N A T L K

973 CAGATAGCTAGCAAAATTAAGAGAACAAATTTGGAAATAATAAAACAATAATCTTTAAGCAATCCTCAGGAGGG
325 Q I A S K L R E Q F G N N K T I I F K Q S S G G

```



24/42

## FIGURE 10C

```

1045 GACCCAGAAATTGTAACGCACAGTTTAAATTGGAGGGGAATTTTCTACTGTAAATTCACACAACTGTTT
349 D P E I V T H S F N C G G E F F Y C N S T Q L F

1117 AATAGTACTTGGTTTAAATAGTACTTGGAGTACTGAAGGGTCAATAACACTGAAGGAAGTGACACAATCACA
373 N S T W F N S T W S T E G S N N T E G S D T I T

1189 CTCCCATGCAGATAAAACAATTATATAACATGGTGCAGGAAGTAGGAAAAGCAATGTATGCCCTCCCATC
397 L P C R I K Q F I N M V Q E V G K A M Y A P P I

1261 AGCGGACAAATTAGATGTTTCATCAATAATTACAGGGCTGCTATTAAACAAGAGATGGTGGTAATAACAACAAT
421 S G Q I R C S S N I T G L L L T R D G G N N N N

1333 GGTCCGAGATCTTCAGACCCTGGAGGAGGAGATATGAGGGACAATTGGAGAGTGAATTATATAATAATAA
445 G S E I F R P G G G D M R D N W R S E L Y K Y K

1405 GTAGTAAATAATTGAACCATTAGGAGTAGCACCACCAAGGCAAGAGAGAGTGGTGCAGAGAGAAAAATGA
469 V V K I E P L G V A P T K A K R R V V Q R E K -
      NotI
1447 GGGGGGGC

```

FIGURE 11A

|            |
|------------|
| FIGURE 11A |
| FIGURE 11B |
| FIGURE 11C |

JR-EL ΔV3-CD4<sup>-</sup>  
1  
1

19  
7

79  
27

139  
47

199  
67

259  
87

319  
107

379  
127

ATGGATGCAATGAAGAGA  
M D A M K R

GGGCTCTGCTGTGCTGTGTGGAGCAGTCTTCGTTTCCGCCAGCCAGGAATC  
G L C C V L L L C G A V F V S P S Q E I

CATGCCCGATTGAGAGAGCGCGCAGAGTAGAAAGTTGTGGGTACAGTCTATTATGGG  
H A R F R R G A R V E K L W V T V Y Y G

GTACCTGTGTGGAAGAAGCAACCACCCTCTATTTTGTGCATCAGATGCTAAAGCATAT  
V P V W K E A T T T L F C A S D A K A Y

GATACAGAGGTACATAATGTTTGGGCCACACATGCCCTGTGTACCCACAGACCCCAACCCA  
D T E V H N V W A T H A C V P T D P N P

CAAGAAGTAGTATTGGAATAATGTAACAGAACATTTTAACATGTGGAAAAATAACATGGTA  
Q E V V L E N V T E H F N M W K N N M V

GAACAGATGCAGGAGGATATAATCAGTTTATGGGATCAAAGCCCTAAAGCCATGTGTAAA  
E Q M Q E D I I S L W D Q S L K P C V K

TTAACCCCACTCTGTGTACTTTAAATTGCAAGGATGTGAATGCTACTAATACCCTAAT  
L T P L C V T L N C K D V N A T N T T N

NarI

▲ Signal cleavage

26/42

## FIGURE 11B

439 GATAGCGAGGGAACGATGGAGAGAGGAGAAATAAAAACTGCTCTTTCAATATCACCACA  
147 D S E G T M E R G E I K N C S F N I T T  
499 AGCATAAGAGATGAGGTGCAGAAAGAATATGCTCTTTTTTATAAACTTGATGTAGTACCA  
167 S I R D E V Q K E Y A L F Y K L D V V P  
559 ATAGATAATAATACCAGCTATAGTTGATAAGTTGTGACACCTCAGTCATTACACAG  
187 I D N N N T S Y R L I S C D T S V I T Q  
619 GCCTGTCCAAAGATATCCTTTGAGCCCAATCCCATACATTATTGTGCCCGCTGTTTT  
207 A C P K I S F E P I P I H Y C A P A G F  
679 GCGATTCTAAAGTGTAATGATAAGACGTTCAATGGAAAGGACCATTGTAAAAATGTCAGC  
227 A I L K C N D K T F N G K G P C K N V S  
739 ACAGTACAATGTACACATGGAATTAGGCCAGTAGTATCAACTCACTGCTGCTAAATGGC  
247 T V Q C T H G I R P V V S T Q L L L N G  
799 AGCTAGCAGAGAAGAGGTAGTAATTAGATCTGACAAATTCACGAAACAATGCTAAACC  
267 S L A E E E V V I R S D N F T N N A K T  
859 ATAATAGTACAGCTGAAGAATCTGTAGAAATTAATTGTACAGGTGCTGGACATTGTAAC  
287 I I V Q L K E S V E I N C T G A G H C N  
919 ATTAGTAGAGCAAAATGGAATGACACTTTAAACAGATAGTTATAAAATTAAGAGAACAA  
307 I S R A K W N D T L K Q I V I K L R E Q

27/42

## FIGURE 11C

979 TTTGAGAAATAACAATAGTCTTTTAATCACTCCTCAGGAGGGAGCCAGAAATTGTAATG  
 327 F E N K T I V F N H S S G G D P E I V M  
  
 1039 CACAGTTTTAATTGTGGAGGAGAAATTTTCTACTGTAAATTCACACACACTGTTTAATAGT  
 347 H S F N C G G E F F Y C N S T Q L F N S  
  
 1099 ACTTGGAAATAATACTGAAGGGTCAATAACACTGAAGGAAATATCATCACACTCCCA  
 367 T W N N N T E G S N N T E G N T I T L P  
  
 1159 TGCAGAATAAAACAAATTATAAATCATGGTGCAGGAAGTAGGAAAGCAATGTATGCCCT  
 387 C R I K Q I I N M V Q E V G K A M Y A P  
  
 1219 CCCATCAGAGGACAAATTAGATGTTTCATCAATAATTACAGGGCTGCTATTAAACAAGAGAT  
 407 P I R G Q I R C S S N I T G L L L T R D  
  
 1279 GGTEGTATTAAATGAGAAATGGGACCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGGAC  
 427 G G I N E N G T E I F R P G G G D M R D  
  
 1339 AATTGGAGAAGTGAATTATATAATAATAAGTAGTAATAATTGAACCATTAGGAGTAGCA  
 447 N W R S E L Y K Y K V V K I E P L G V A  
  
 1399 CCCACCAAGGCAAGAGAAGAGTGGTGCAAAGAGAGAAATGAGCGGCGGC  
 487 P T K A K R R V V Q R E K -

NotI

28/42

## FIGURE 12A

|            |
|------------|
| FIGURE 12A |
| FIGURE 12B |
| FIGURE 12C |

LAI CD4<sup>-</sup>

1 tPA signal sequence  
 1 ATGGATGCAATGAAGAGAGGGCTCTGCTGT  
 M D A M K R G L C C

37 GTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCAGCCAGGAAATCCATGCCCGATTTCAGAAAGAGCGCC  
 13 V L L L C G A V F V S P S Q E I H A R F R R G A  
 Nal

109 AGACAGAAATTTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAGGAGCAACCACTCTATT  
 37 R T E K L W V T V Y Y G V P V W K E A T T T L F  
 ▲ Signal cleavage

181 TGTGCATCAGATGCTAAAGCATAATGATACAGAGGTACATAATGTTTGGGCCACACATGCCCTGTGTACCCACA  
 61 C A S D A K A Y D T E V H N V W A T H A C V P T

253 GACCCCAACCCACAAGTAGTATTGGTAATGTGACAGAAATTTTAAATGTGGAAATGACATGGTA  
 85 D P N P Q E V V L V N V T E N F N M W K N D M V

325 GAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAGCCCTAAAGCCATGTGTAAATTAACCCCACTC  
 109 E Q M H E D I I S L W D Q S L K P C V K L T P L

397 TGTGTTAGTTTAAAGTGCACCTGATTGGGGGAATGCTACTAATACCAATAGTAGTAGTAGC  
 133 C V S L K C T D L G N A T N T N S S N T N S S S

29/42

## FIGURE 12B

469 GGGGAATGATGGAGAAAGGAGAGATAAAAACTGCTCTTTCAATATCAGCACAAAGCATAAGAGGTAAG  
157 G E M M E K G E I K N C S F N I S T S I R G K

541 GTGCAGAAAGAATATGCATTTTTATATAAACTTGATATAATACCAATAGATAATGATACTACCAGCTATACG  
181 V Q K E Y A F F Y K L D I I P I D N D T T S Y T

613 TTGACAAGTTGTACACCTCAGTCATTACACAGGCCCTGTCCAAGGTATCCTTTGAGCCCAATTCCCATACAT  
205 L T S C N T S V I T Q A C P K V S F E P I P I H

685 TATTGTGCCCCGGCTGGTTTTCGATTCTAAAATGTAATAAAGACGTTCAATGGAACAGGACCATGTACA  
229 Y C A P A G F A I L K C N N K T F N G T G P C T

757 AATGTCAGCACAGTACAATGTACACATGGAAATTAGGCCAGTAGTATCAACTCAACTGCTGTGTAATGCGAGT  
253 N V S T V Q C T H G I R P V V S T Q L L L N G S

829 CTAGCAGAAGAAGGAGTAGTAATTAGATCTGCCAATTTCCACAGACAATGCTAAACCATAATAGTACAGCTG  
277 L A E E E V V I R S A N F T D N A K T I I V Q L

901 AACCAATCTGTAGAAATTAATTGTACAGACCCCAACAATAACAAGAAAGTATCCGTATCCAGAGGGGA  
301 N Q S V E I N C T R P N N N T R K S I R I Q R G

973 CCAGGGAGAGCATTGTGTACAAATAGGAAAAATAGGAAATATGAGACAAGCACATTGTAAACATTAGTAGGCA  
325 P G R A F V T I G K I G N M R Q A H C N I S R A

30/42

## FIGURE 12C

1045 AAATGGAATGCCACTTTAAACAGATAGCTAGCAAAATTAAGAGAACAAATTGGAAATAATAAACAAATAATC  
 349 K W N A T L K Q I A S K L R E Q F G N N K T I I  
  
 1117 TTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTACGCACAGTTTAAATTGTGGAGGGGAATTTTCTAC  
 373 F K Q S S G G D P E I V T H S F N C G G E F F Y  
  
 1189 TGTAAATCAACACAACCTGTTTAATAGTACTTGGTTTAATAGTACTTGGAGTACTGAAGGGTCAATAACACT  
 397 C N S T Q L F N S T W F N S T W S T E G S N N T  
  
 1261 GAAGGAAGTGACACAATCACACTCCCATGCAGAAATAAAACAATTATATAACATGGTGAGGAAGTAGGAAAA  
 421 E G S D T I T L P C R I K Q F I N M V Q E V G K  
  
 1333 GCAATGTATGCCCTCCCATCAGCGGACAAATTAGATGTTTCATCAATAATTACAGGGCTGTATTAAACAAGA  
 445 A M Y A P P I S G Q I R C S S N I T G L L L T R  
  
 1405 GATGGTGGTAATAACAACAATGGGTCCGAGATCTTCAGACCTGGAGGAGGAGATATGAGGACAATTGGAGA  
 469 D G G N N N N G S E I F R P G G G D M R D N W R  
  
 1477 AGTGAATTATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAGAGAAGA  
 493 S E L Y K Y K V V K I E P L G V A P T K A K R R  
  
 1549 GTGGTGCAGAGAGAAAATGAGCGGCGGC  
 517 V V Q R E K -

NotI

FIGURE 13A

|            |
|------------|
| FIGURE 13A |
| FIGURE 13B |
| FIGURE 13C |
| FIGURE 13D |

JR-FL CD4'

|     |  |  |
|-----|--|--|
| 1   | ATGGATGCAATGAAGAGA   |  |
| 1   | M D A M K R  |  |
| 19  | GGGCTCTGCTGTGCTGTGCTGTGGAGCAGTCTTCGTTTCGCCAGCCAGGAATC        |  |
| 7   | G L C C V L L L C G A V F V S P S Q E I                      |  |
|     | NarI   |  |
| 79  | CATGCCCGATTGAGAGAGGGCGGAGAGTAGAAAGTTGTGGGTACAGTCTATTATGGG    |  |
| 27  | H A R F R R G A R V E K L N V T V Y Y G                      |  |
|     | ▲ Signal cleavage  |  |
| 139 | GTACCTGTGTGGAAGAACCAACCACCTCTATTTTGTGCATCAGATGCTAAAGCATAT    |  |
| 47  | V P V W K E A T T T L F C A S D A K A Y                      |  |
| 199 | GTACAGAGGTACATAATGTTTGGGCCACACATGCCCTGTGTACCCACAGACCCCAACCCA |  |
| 67  | D T E V H N V W A T H A C V P T D P N P                      |  |
| 259 | CAAGAAGTAGTATTGGAATAATGTAACAGACATTTTAACATGTGGAAATAACATGGTA   |  |
| 87  | Q E V V L E N V T E H F N M W K N N M V                      |  |



32/42

## FIGURE 13B

319 GAACAGATGCAGGAGGATATAATCAGTTTATGGGATCAAGCCTAAAGCCATGTGTAAAA  
107 E Q M Q E D I I S L W D Q S L K P C V K

379 TTAACCCCACTCTGTGTACTTTTAATTGCAAGGATGTGAATGCTACTAATACCCTAAT  
127 L T P L C V T L N C K D V N A T N T T N

439 GATAGCGAGGGAACGATGGAGAGAGGAGAAATAAAACTGCTCTTTCAATATCACCACA  
147 D S E G T M E R G E I K N C S F N I T T

499 AGCATAAGAGATGAGGTGCAGAAAGAATATGCTCTTTTATAAACTTGATGTAGTACCA  
167 S I R D E V Q K E Y A L F Y K L D V V P

559 ATAGATAATAATAACCAAGCTATAGTTGATAAGTTGTGACACCTCAGTCATTACACAG  
187 I D N N N T S Y R L I S C D T S V I T Q

619 GCCTGTCCAAGATATCCTTTGAGCCCAATCCCATACATTATTGTGCCCGGCTGGTTT  
207 A C P K I S F E P I P I H Y C A P A G F

33/42

## FIGURE 13C

|     |   |
|-----|---|
| 679 | CGATTCTAAAGTGTAATGATAAGACGTTCAATGGAAAGGACCATGTAAATGTCAGC      |
| 227 | A I L K C N D K T F N G K G P C K N V S                       |
| 739 | ACAGTACAATGTACACATGGAAATTAGGCCAGTAGTATCAACTCAACTGCTGCTAAATGGC |
| 247 | T V Q C T H G I R P V V S T Q L L L N G                       |
| 799 | AGTCTAGCAGAAGAGGAGTAGTAATTAGATCTGACAATTTCACGAACAATGCTAAACCC   |
| 267 | S L A E E V I R S D N F T N N A K T                           |
| 859 | ATAATAGTACAGCTGAAGAATCTGTAGAAATTAATTGTACAAGACCCCAACAATACA     |
| 287 | I I V Q L K E S V E I N C T R P N N T                         |
| 919 | AGLAAAGTATACATATAGGACCAGGAGAGCATTTTATACAGGAGAAATAATAGGA       |
| 307 | R K S I H I G P G R A F Y T T G E I I G                       |
| 979 | GATATAAGACAAGCACATTGTACATTAGTAGAGCAAAATGGAAATGACACTTTAAACAG   |
| 327 | D I R Q A H C N I S R A K W N D T L K Q                       |

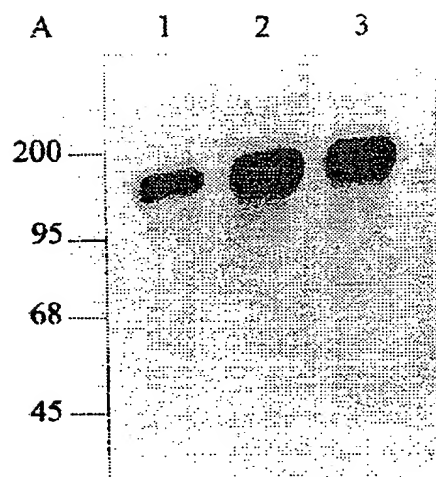
34/42

## FIGURE 13D

|      |  |
|------|--|
| 1039 | ATAGTTATAAAATTAGAGAACAAATTGAGAAATAAACAATAGTCTTTAATCACTCCTCA    |
| 347  | I V I K L R E Q F E N K T I V F N H S S                        |
| 1099 | GGAGGGACCCAGAAATTGTAATGCACAGTTTAAATTGTGGAGGAGAAATTTTCTACTGT    |
| 367  | G G D P E I V M H S F N C G G E F F Y C                        |
| 1159 | AATTCAACACAACACTGTTTAAATAGTACTTGGAAATAATACTGAAGGGTCAATAACACT   |
| 387  | N S T Q L F N S T W N N N T E G S N N T                        |
| 1219 | GAAGGAAATACATCACTCCCATGCAGAAATAAACAATAATTATAACATGGTGCAGGAA     |
| 407  | E G N T I T L P C R I K Q I I N M V Q E                        |
| 1279 | GTAGGAAAGCAATGTATGCCCTCCCATCAGAGGACAATAATTAGATGTTTCATCAATATT   |
| 427  | V G K A M Y A P P I R G Q I R C S S N I                        |
| 1339 | ACAGGGCTGCTATTAAACAAGAGATGGTGGTATTAAATGAGAAATGGACCGAGATCTTCAGA |
| 447  | T G L L L T R D G G I N E N G T E I F R                        |
| 1399 | CCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAATAAGTAGTA       |
| 467  | P G G G D M R D N W R S E L Y K Y K V V                        |
| 1459 | AAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAGAGAGAGAGTGGTGCAAGAGAA     |
| 487  | K I E P L G V A P T K A K R R V V Q R E                        |
|      | NotI   |
| 1519 | AAATGAGCGGCGGC   |
| 507  | K  |

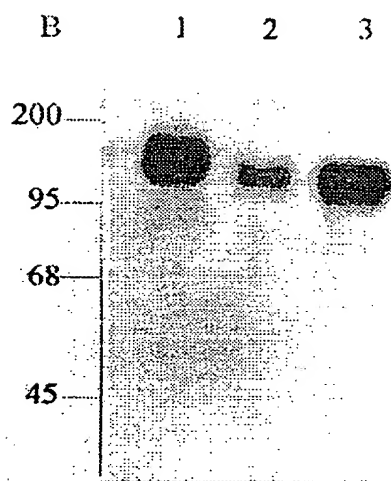
35/42

FIGURE 14A



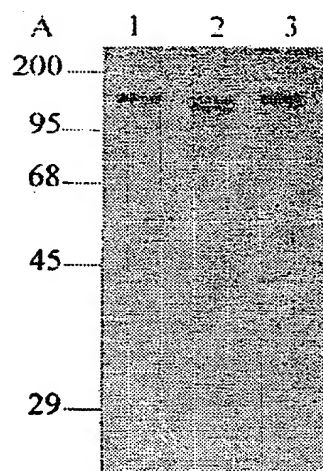
36/42

FIGURE 14B



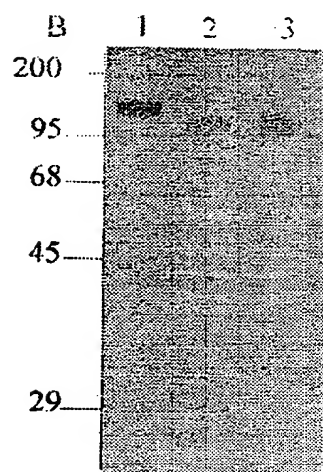
37/42

FIGURE 15A



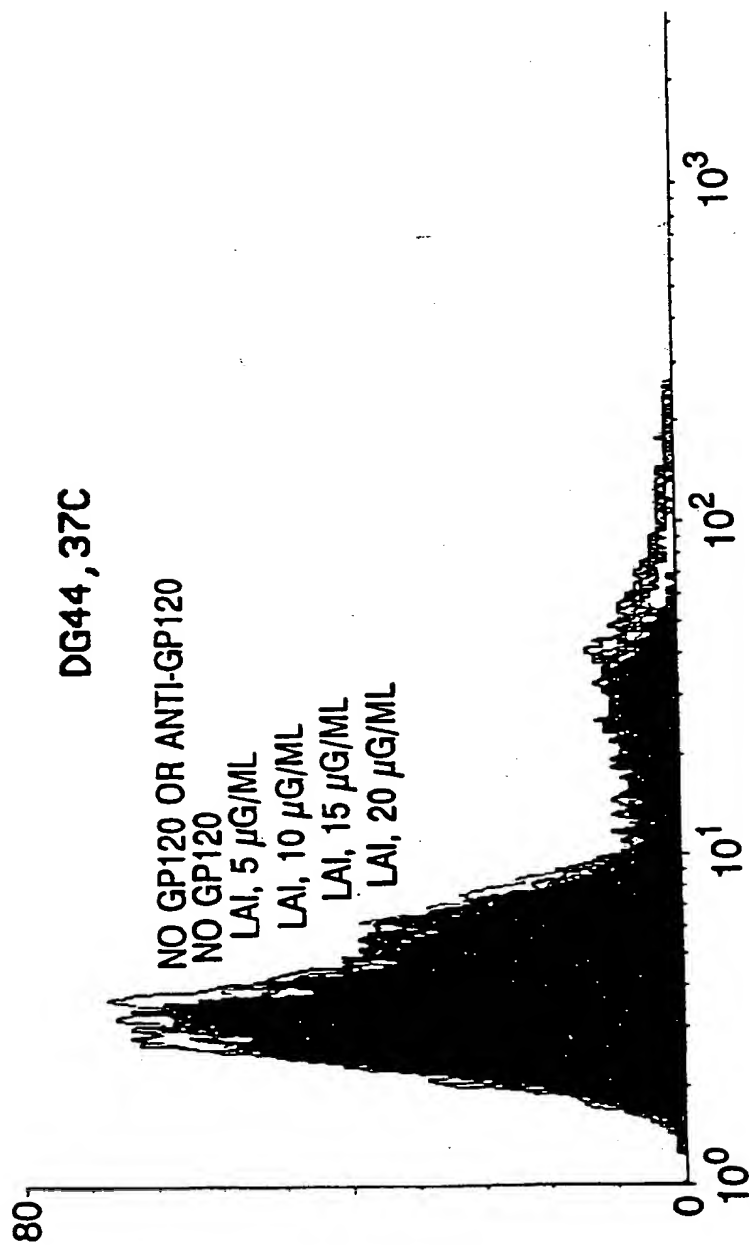
38/42

FIGURE 15B



39/42

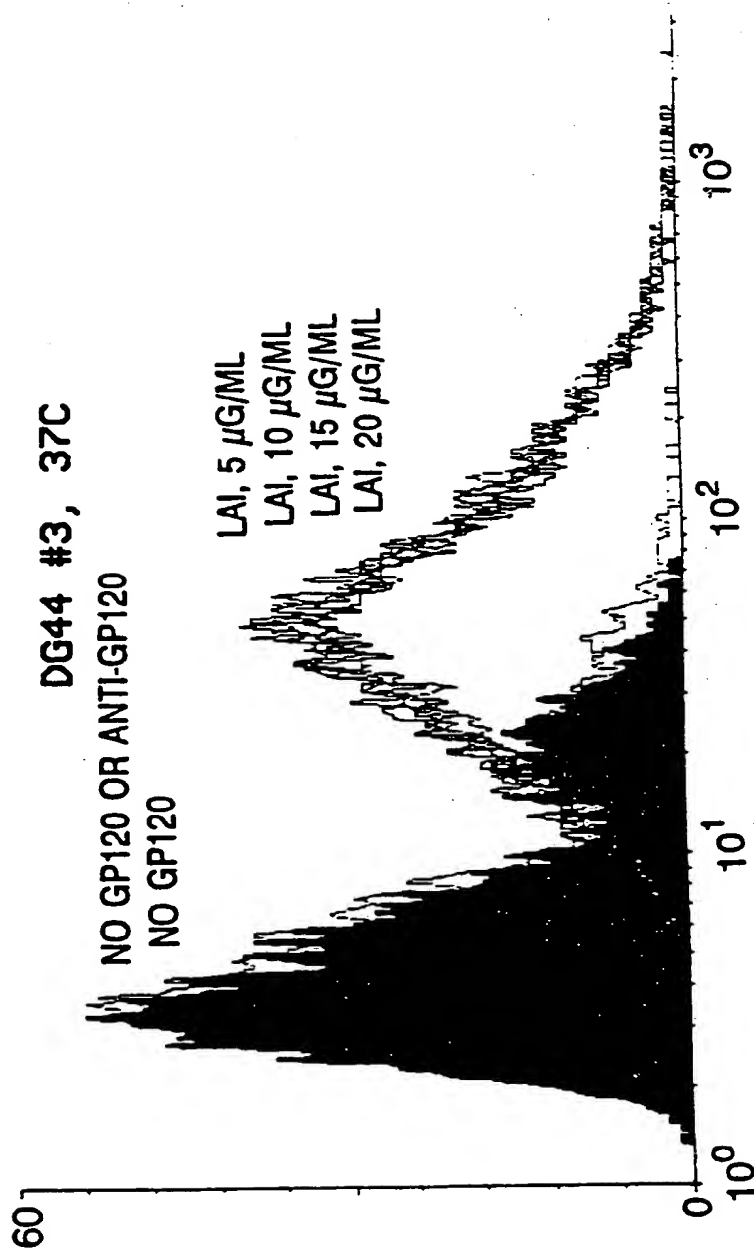
FIGURE 16A





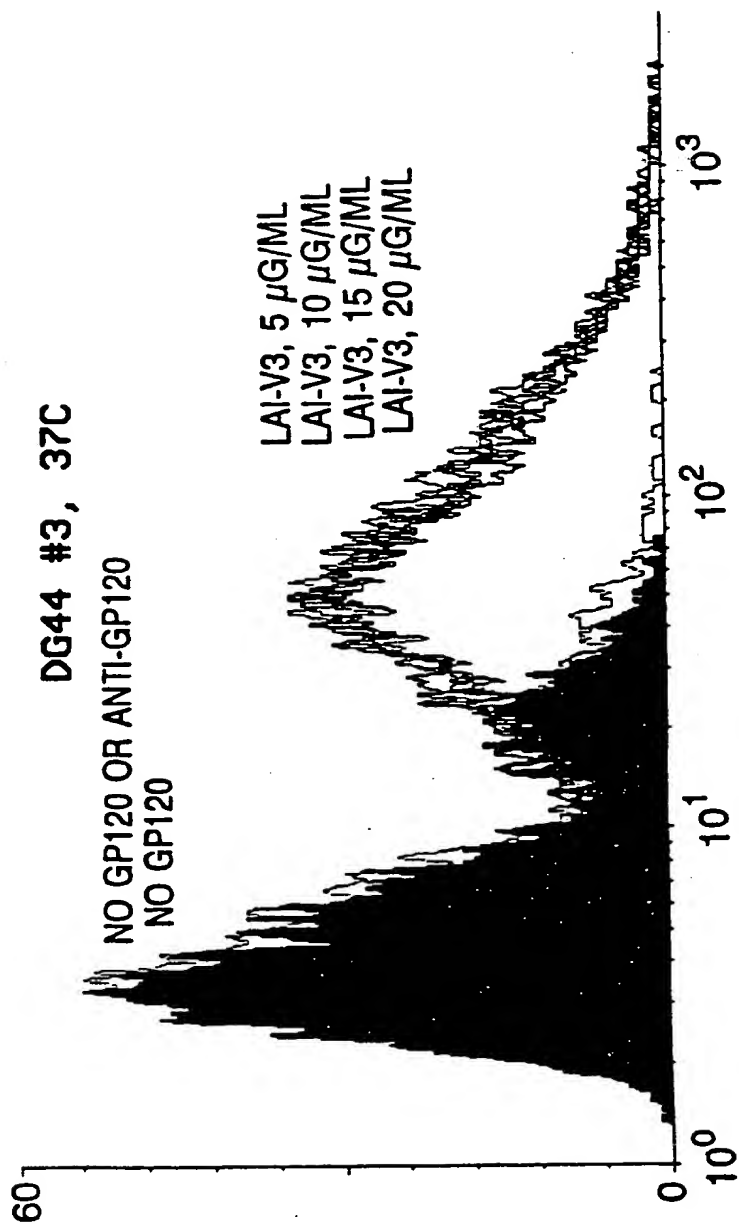
40/42

FIGURE 16B



41/42

FIGURE 16C

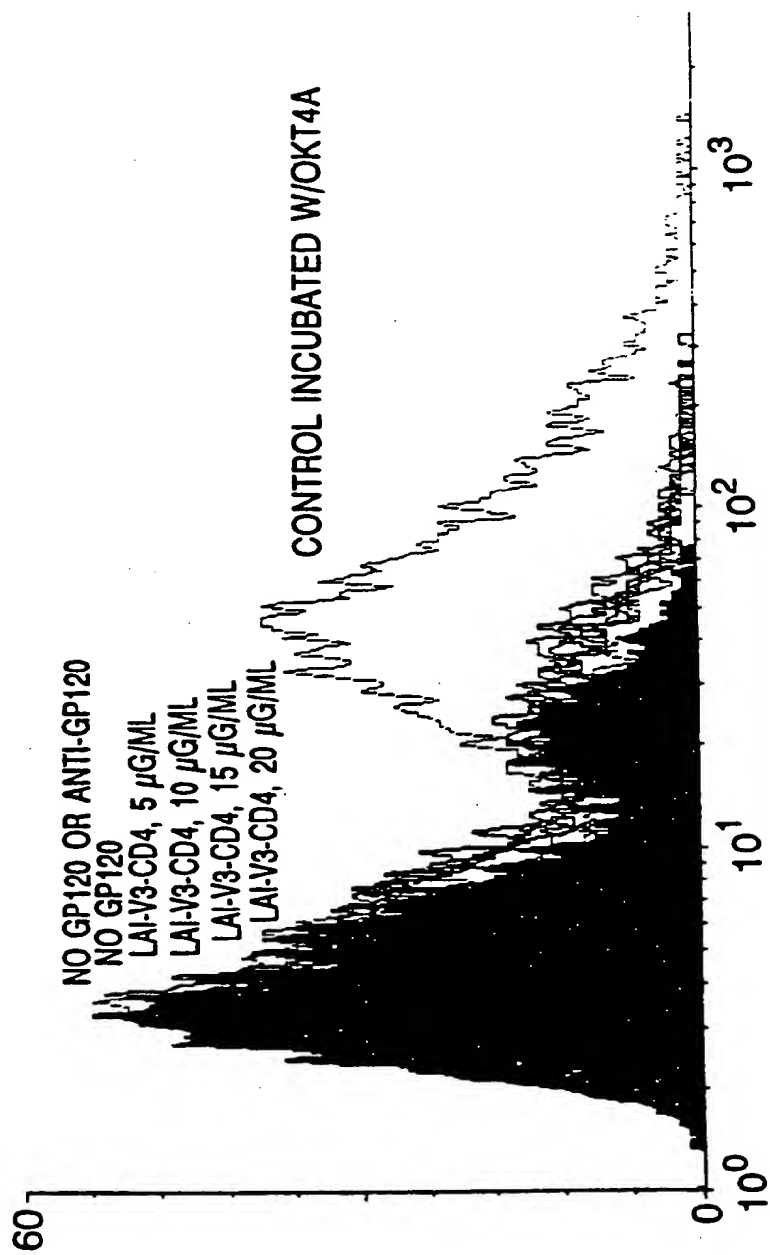


SUBSTITUTE SHEET (RULE 26)

42/42

FIGURE 16D

DG44 #3, 37C



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US94/03282

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) : Please See Extra Sheet.

US CL : 424/88, 89; 536/27; 530/395

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/88, 89; 536/27; 530/395

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Dialog, search terms: HIV-1, mutation, V3 loop, C4 region, envelope glycoprotein, vaccines, nucleic acid

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No. |
|-----------|--|-----------------------|
| Y         | Science, Volume 252, issued 17 May 1991, S. Wain-Hobson, et al, "LAV Revisited: Origins of the Early HIV-1 Isolates from Institut Pasteur", pages 961-965, see entire article. | 6,7                   |
| Y         | US, A, 5,030,449 (BERZOFKY ET AL) 09 July 1991, cols. 4-7.   | 1-27                  |
| Y         | WO, A, 91/15512 (GREGORY ET AL) 17 October 1991, entire patent.  | 1-27                  |
| Y         | WO, A, 91/11461 (PASEK ET AL) 08 August 1991, see entire patent.   | 1-27                  |

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

|          |   |   |
|----------|---|---|
| *<br>"A" | Special categories of cited documents:<br>document defining the general state of the art which is not considered to be of particular relevance                      | "T"<br>later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  |
| "E"      | earlier document published on or after the international filing date  | "X"<br>document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone   |
| "L"      | document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "Y"<br>document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "O"      | document referring to an oral disclosure, use, exhibition or other means  | "&"<br>document member of the same patent family  |
| "P"      | document published prior to the international filing date but later than the priority date claimed  |   |

Date of the actual completion of the international search

09 JUNE 1994

Date of mailing of the international search report

JUN 24 1994

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

LYNETTE F. SMITH

Telephone No. (703) 308-0196

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US94/03282

## A. CLASSIFICATION OF SUBJECT MATTER:

IPC (5):

A61K 39/12, 39/00; C07K 17/00, 3/00, 13/00, 15/00; C07H 15/12, 13/00